Set Book

The Open University Press

Personality and Learning 1

9599

Personality and Learning 1

STATE INSTITUTE OF EDUCATION
Govt of West Bengal

The Open University
Faculty of Educational Studies

The following people were responsible for the selection of articles in this Reader:
Asher Cashdan (Chairman of the Course Team)
Ann Floyd
Paul Kline (Consultant)
Victor Lee
Larry Litt
John Oates
Roy Webberley
Joan M. Whitehead
Jane Wolfson

Other members of the Personality and Learning Course Team include:
Richard Argent
Beryl Crooks
Sheila Dale
Rhiannon Davies
Arthur James
Edward Milner
Caroline Morrow Brown
Derek Rowntree
Meg Sheffield
Sue Vaudin

Editor: Joan M. Whitehead

Personality and Learning 1

A Reader prepared by the Personality and Learning Course Team at The Open University

Edited for the Course Team by
Joan M. Whitehead

STATE INSTITUTE OF EDUCATION
Govt. of West Bengal

HODDER AND STOUGHTON

in association with
The Open University Press

[illegible] 19.7.93
[illegible] 7079

137
WHI

~~9599~~

ISBN 0 340 20472 9 Paper
ISBN 0 340 20473 7 Boards

Selection and editorial material
copyright © The Open University 1975

All rights reserved. No part of this publication may be
reproduced or transmitted in any form or by any means,
electronic or mechanical, including photocopy, recording,
or any information storage and retrieval system,
without permission in writing from the publisher.

Printed and bound in Great Britain for
Hodder and Stoughton Educational,
a division of Hodder and Stoughton Ltd,
St Paul's House, Warwick Lane, London EC4P 4AH,
by Hazell Watson & Viney Ltd, Aylesbury, Bucks.
Computer Typesetting by Print Origination, Bootle, Merseyside

Contents

STATE INSTITUTE OF EDUCATION
Govt. of West Bengal

Section One Introduction

Personality in the learning situation *Asher Cashdan* 1

1 Sources of information *T.R. Sarbin and W.C. Coe* 4

2 Reading research articles *T.R. Sarbin and W.C. Coe* 13

3 Nursery teachers' perceptions of their pupils: an exploratory study *B. Thompson* 17

Section Two Personality theories and dimensions

Introduction *Asher Cashdan* 35

4 An outline of psycho-analysis (chapters 1 to 5) *S. Freud* 37

5 Empirical studies of the Oedipus and castration complexes *P. Kline* 56

6 The validity of Gottheil's Oral Trait Scale in Great Britain *P. Kline* 73

7 Personality testing and the school child *R.B. Cattell* 78

8 Primaries or second-order factors: a critical consideration of Cattell's 16PF battery *H.J. Eysenck* 90

9 Personality and academic attainment *N.J. Entwistle* 96

Section Three Motivation and learning

Introduction *Joan M. Whitehead* 115

10 Experience and the development of motivation: some reinterpretations *J.McV. Hunt* 116

11 The will to learn *J.S. Bruner* 132

12 Social class and achievement motivation *D.F. Swift* 142

13 The science of learning and the art of teaching *B.F. Skinner* 158

14 The treatment of socially oriented underachievers: a case study 163
B. Schaeffer, A. Harris and J.M. Greenbaum
15 The effects of four programs of classroom intervention on the intellectual and language development of 4-year-old disadvantaged children 168
M.B. Karnes, J.A. Teska and A.S. Hodgins

Section Four The course of development

Introduction *John Oates* 189
16 Self-conceptions, motivations and interpersonal attitudes of late- and early-maturing boys 191
P.H. Mussen and M.C. Jones
17 On equivalence 204
R.R. Olver and J. Rigney Hornsby
18 Provoked correspondence and equivalence of corresponding sets 221
J. Piaget and A. Szeminska
19 The understanding of invariance by very young children 237
P.E. Bryant

Section Five Cognitive styles

Introduction *Ann Floyd* 255
20 Learning strategies and individual competence 257
G. Pask and B.C.E. Scott
21 Some implications of research on cognitive style for problems of education 288
H.A. Witkin
22 Conceptual impulsivity and inductive reasoning 315
J. Kagan, L. Pearson and L. Welch

Section Six Creativity and intelligence

Introduction *Victor Lee* 327
23 Lost: our intelligence? Why? 329
Q. McNemar
24 Race, intelligence and genetics: the differences are real 345
A.R. Jensen
25 Environment: the cumulation of effects is yet to be understood 357
A.L. Stinchcombe
26 The person, the product and the response: conceptual problems in the assessment of creativity 368
P.W. Jackson and S. Messick
27 Creativity and curriculum structure 384
E. Ogilvie

Index 395

Acknowledgments

The editor and publisher wish to thank the following for permission to reprint the articles included in this book:

Schenkman Publishing Company for 'Sources of information' and 'Reading research articles' (reprinted from *The Student Psychologist's Handbook: A Guide to Sources* by T. R. Sarbin and W. C. Coe, 1968, pp. 25–36, 47–51); Miss B. L. Thompson for 'Nursery teachers' perception of their pupils: an exploratory study' (first published in this volume); Sigmund Freud Copyrights Limited, The Institute of Psycho-Analysis and The Hogarth Press Limited for 'An outline of psycho-analysis' (reprinted from Volume XXIII of *The Standard Edition of the Complete Psychological Works of Sigmund Freud,* revised and edited by James Strachey, 1964, pp. 144–71); Associated Book Publishers Limited/Methuen Educational Limited for 'Empirical studies of the Oedipus and castration complexes' (reprinted from *Fact and Fantasy in Freudian Theory* by Paul Kline, 1972, pp. 97–110); The Journal Editor and Paul Kline for 'The validity of Gottheil's Oral Trait Scale in Great Britain' (reprinted from *Journal of Personality Assessment,* 1973, *37,* pp. 551–4); Penguin Books Limited for 'Personality testing and the school child' (reprinted from *The Scientific Basis of Personality* by Raymond B. Cattell, 1965, pp. 309–25. Copyright © Raymond B. Cattell, 1965); Cambridge University Press for 'Primaries or second-order factors: a critical consideration of Cattell's 16PF battery' by H. J. Eysenck (reprinted from *British Journal of Social and Clinical Psychology,* 1972, *11,* pp. 265–9); Scottish Academic Press Limited for 'Personality and academic attainment' by N. J. Entwistle (reprinted from *British Journal of Educational Psychology,* 1972, *4,* pp 137–51); The Society for Research in Child Development, Inc., for 'Experience and the development of motivation: some reinterpretations' by J. McV. Hunt (reprinted from *Child Development,* 1960, *31,* pp. 489–504), 'Self-conceptions, motivations and interpersonal attitudes of late- and early-maturing boys' by P. H. Musson and M. C. Jones (reprinted from *Child Development,* 1957, *28,* pp. 243–56), and 'Conceptual impulsivity and inductive reasoning' by J. Kagan, L. Pearson and L. Welch (reprinted from *Child Development,* 1966, *37,* pp. 583–94); NFER Publishing Company Limited for 'Social class and achievement motivation' by D. F. Swift (reprinted from *Educational Research,* 1966, *8,* pp. 83–95) and 'Creativity and curriculum structure' by E. Ogilvie (reprinted from *Educational Research,* 1974, *16,* pp. 126–32); The Belknap Press of Harvard University Press, Cambridge, Mass., for 'The will to learn' (reprinted from *Toward a Theory of Instruction* by Jerome S. Bruner, 1966, pp. 113–28. Copyright © 1966 by the President and Fellows of Harvard College); Harvard Educational Review for 'The science of learning and the art of teaching' (reprinted from *The Technology of Teaching* by B. F. Skinner, 1968, pp. 14–21, Appleton-Century-Crofts) and 'Environment: the cumulation of effects is yet to be understood' by A. L. Stinchcombe (reprinted from *Harvard Educational Review,* 1969, *39,* pp. 511–22); Behavioural Publications, a subsidiary of Human Sciences, Inc., 72 Fifth Avenue, New York, N.Y. 10011, for 'The treatment of socially oriented underachievers: a case study' by B. Schaeffer, A. Harris and J. M. Greenbaum (reprinted from *Journal of School Psychology,* 1968-69, 7, 4, pp. 70–3); American Orthopsychiatric Association, Inc., for 'The effects of four programs of classroom intervention on the intellectual and language development of 4-year-old disadvantaged children' by M. B. Karnes, J. A. Teska and A. S. Hodgins (reprinted from *American Journal of Orthopsychiatry,* 1970, *40,* 1, pp. 58–76. Copyright © 1970 the American Orthopsychiatric Association, Inc. Reproduced by permission); Canadian Psychological Association for 'The understanding of invariance by very young children' by P. E. Bryant (reprinted from *Canadian Journal of Psychology/Review of Canadian Psychology,* 1972, *26,* pp. 78–96. Copyright 1972, Canadian Psychological Association. Reprinted by permission); John Wiley and Sons, Inc., for 'On equivalence' by R. R. Olver and J. R. Hornsby (reprinted from *Studies in Cognitive Growth* by J. S. Bruner *et al,* 1966, pp. 68–85. Copyright © 1966. Reprinted by permission of John Wiley and Sons, Inc.); Routledge and Kegan Paul Limited for 'Provoked correspondence and equivalence of corresponding sets' by J. Piaget and A. Szeminska (reprinted from *The Child's Concept of Number* by Jean Piaget, 1952, pp. 41–56); Dr Gordon Pask for 'Learning strategies and individual competence' by G. Pask and B. C. E. Scott (reprinted in abridged form from *International Journal of Man-Machine Studies,* 1972, *4,* 217-53); Grune and Stratton, Inc., and the author for 'Some implications of research on cognitive style for problems of education' by H. A. Witkin (reprinted from *Professional School Psychology,* Vol. 3, edited by M. B. Gottesgen and G. B. Gottesgen, 1969, pp. 198–227. Reprinted by permission); American Psychological Association for 'Lost: our intelligence? Why?' by Q. McNemar (reprinted from *American Psychologist,* 1964, *19,* pp. 871–82. Copyright 1964 by the American Psychological Association. Reprinted by permission); *Psychology Today* magazine for 'Race, intelligence and genetics: the differences are real' by A. R. Jensen (reprinted from *Psychology Today,* 1973, *7,* pp. 80–6. Copyright © 1973 Ziff-Davis Publishing Company. All rights reserved); Duke University Press for 'The person, the product and the response: conceptual problems in the assessment of creativity' by P. W. Jackson and S. Messick (reprinted from *Journal of Personality,* 1965, *33,* pp. 1–19, by permission of the Publisher. Copyright 1965, Duke University Press, Durham, North Carolina).

Section One
Introduction

Personality in the learning situation

Learning, to the psychologist, is not confined to the classroom or lecture-hall, but is used in the broader sense of an ongoing process of adaptation to the environment. The infant starts to learn as soon as he is born; in fact, some learning occurs even in the pre-natal period. Throughout early infancy he is learning social and intellectual skills from his environment. By the time the child reaches school he is well on the way to full command of his native language and has developed some powers of thinking and reasoning. He continues to learn, both in and out of the classroom, for the rest of his life.

But individuals vary greatly in their rate of learning and in what they learn. So it is natural that educational psychologists have spent a great deal of energy in studying the reasons for individual differences and in trying to improve people's learning ability and efficiency. Although learning has been studied in many ways, the educational psychologist has tended to concentrate on studying abilities and their structure, and on designing measures of intelligence and attainment. But in recent years it has become steadily more apparent that the child's attainment, even in tests of 'pure' intelligence, depends to quite a considerable extent on the experiences he has enjoyed. One of the most seminal books in the last twenty years has been Hunt's *Intelligence and Experience*, published in 1961, in which he discusses in detail the role of experience in intellectual development, taking into account both traditional studies and the more recent work of the Swiss psychologist, Piaget.

Side by side with such developments, but not often well integrated with them, there has been an interest in the field of personality, its structure and measurement. This has often been pursued without direct reference to the intellectual field. In other words, differences in personality characteristics, or in the ways in which people make and sustain relationships with each other, have been examined in relative isolation from the effects of these differences on the individual's ability, and on the likelihood of his being an effective learner. Personality in this restricted sense is presented in Section 2 of this Reader, though even there we are aware of the need to relate it to educational issues. But if, as in the whole of the course to which this volume acts as a companion, we interpret the term at its widest, then it must include all those characteristics we have already mentioned. That is to say, the individual's abilities, his experiences and their effects upon him are all as much a part of his personality as it develops as are those characteristics we more immediately recognize as part of his personality.

So, in constructing the Open University's course *Personality and Learning*—a course which is in essence a basic introduction to educational and child psychology—our aim has been to examine the individual learner in terms of his total personality. The course is divided into blocks, each devoted to a different topic; in this way we look at our learner from a number of different angles and in a variety of contexts, in the hope that the student may thus build up his own total picture. Some of the blocks are more theoretical, as when we specifically examine theories of learning and of personality, while others are more firmly applied, as in the treatment of classroom interaction, but in all cases some educational implications are discussed. Having said this, we can nevertheless discern a tendency for the course to proceed from theory to practice, so that this first volume of readings covers the more theoretical ground, while the second volume has more of an applied flavour.

Following the three introductory papers, there are ten course blocks. This volume covers five of them: personality theories, motivation and learning, the developmental process, cognitive styles and some consideration of the interlocking areas of intelligence and creativity. In a second volume we go on to explore social relationships, cross-cultural studies, adjustment and learning difficulties and classroom interaction.

The course *Personality and Learning* has been designed for the needs of Open University second level students who have an interest in the applications of psychology to education and child development. We have tried in this Reader to introduce them to the thinking and the experimental work of those psychologists (and, occasionally, other social scientists) whose work has seemed to us most accessible and relevant. Papers are, therefore, presented with a minimum of editorial interference or comment, this being reserved where necessary for the Open University students' correspondence text. Past experience, however, leads us to expect that these volumes will prove useful to a variety of other students of education and psychology, either on their own or used in conjunction with the correspondence texts. The Course Team will be delighted if this proves to be the case. The casual reader need only keep in mind that not all the papers are included as models, nor is the coverage within each topic necessarily completely balanced.

This introduction is followed by three papers. The first two, taken from Sarbin and Coe's *The Student Psychologist's Handbook,* provide some useful guidelines to students approaching this field for the first time. Readers are strongly recommended to tackle these before going on to study any of the subsequent chapters. The section ends with a specially contributed paper by Barbara Thompson, which is included to demonstrate one method of approaching the scientific study of personality. Personality dimensions should grow out of actual usage, rather than being constructed in the theorist's armchair. Both Raymond Cattell and George Kelly have developed their own solution to this problem. Cattell uses the printed dictionary–an exhaustive but exhausting method–while Kelly uses the 'personal' dictionary; that is, he starts from the words which people actually use to describe others and builds the system from there. The Thompson paper illustrates how this may be done with a small sample of nursery teachers who were asked to describe their pupils.

In the second section, as we have said, personality is interpreted rather more narrowly. The readings consist of presentations of Freud's work and that of the factor-analysts, with some discussion of their strengths and limitations. Both are rather extreme positions and intermediate theories are not considered in this reader; though Open University students will also be spending some time considering Murray's approach. But the section provides the reader with some essential background to the broader areas which are to follow.

Section 3 deals with basic research in the fields of motivation and learning, and emphasizes the relationship between them. There is a good balance between the more theoretical contributions of Hunt and Skinner and their application in school situations.

Section 4 is concerned with the developmental aspects of intellectual abilities and the relationship between physical maturation and individuals' self-concepts. A thread that runs through the papers is that children vary in the rate at which they grow, not only physically but also intellectually, and it may be that teachers need to pay attention not just to chronological but also to 'developmental' age.

The fifth section follows this cognitive emphasis by discussing the question of differences not just in abilities, but in preferred styles of learning and their relevance to the educational situation. This is an under-researched field where more knowledge could help enormously towards the individualization of teaching.

Finally, Section 6 considers a major contemporary debate: the nature of intelligence and how far intellectual differences can be attributed to hereditary or environmental causes, or whether this is an unreal question. The latter part of the section looks at the nature of creativity and its relationship to intelligence; again, we are always aware of the need to relate these issues to classroom practice.

Asher Cashdan

Reference

HUNT, J. McV. (1961) *Intelligence and Experience.* New York: Ronald Press.

1 Sources of information

T. R. Sarbin and W. C. Coe

The ultimate worth of a scholarly work depends on the information used in its composition. Two major tasks are involved in obtaining information: locating it, and evaluating it. To locate information, one must know not only *where* to look but also *how* to look for it. In evaluating information, qualitative and quantitative aspects must be taken into account.

The quality of information is determined by considering the methods used in obtaining it. The way in which information is obtained indicates the degree of scientific rigour involved, and consequently, its usefulness to the particular project. The least rigorous sources of information are useful for speculation and for developing ideas, methods of moderate rigour serve to refine ideas and may provide some support for hypotheses, while rigorous experimental procedures provide the most definitive answers.

The quantitative characteristics of information are related to the degree of sophistication of measurement. Generally speaking, the more sophisticated the measurement, the more reliable the information. As quantification becomes more refined, the investigator is better able to qualify his statements; statements may be made about different effects in relation to different degrees of a variable. For instance, information on responsiveness to hypnosis appears in different quantitative levels. One report will state a gross differentiation, '*most* people can be hypnotised', while another report will indicate the percentage of persons who reach different levels of depth as measured by a standard hypnotic scale. Although the two sources provide information on the same topic, it is clear that the information from the latter report is more refined than that of the first. Increasing quantification also increases the possibility that a study can be repeated by other investigators. And the repetition of experiments (when the outcomes are congruent) adds confidence to the validity and reliability of the information.

Observation is the basic source of scientific information. Scientists, indeed most people, base their ideas about the world on what they and others have observed. Many observations are made directly; that is, we record what we perceive through our senses. For instance, an observer may record the different behaviours exhibited by two children playing hopscotch. Other observations are made indirectly with the help of measuring instruments. Body temperature is observed with the aid of a thermometer,

---

Source: SARBIN, T.R. and COE, W.C. (1968) *The Student Psychologist's Handbook: A Guide to Sources.* Cambridge, Massachusetts: Schenkman, 25-36.

although it may also be observed less precisely by placing a hand on the subject's forehead. In psychology, it is common to measure personality characteristics indirectly through questionnaires and other testing devices constructed for this purpose. The psychologist may make inferences about personality characteristics from the subject's response to questions of the type: 'do you ever have nightmares?', 'do you enjoy playing games?', 'do you like to watch television?'

Observations may be placed in two major categories–personal observations, and the observations of others. The remainder of the chapter will be devoted to examining these two sources of information; our emphasis will be, first, how does one obtain the information? and second, how does one evaluate the information?

Personal observations

Several procedures are available for gathering information through personal observation. Although each differs in its qualitative and quantitative features, each may still be useful in planning and preparing a research project. We shall discuss the following kinds of personal observations: casual observation, focused observation, and controlled observation.

I Casual observation

The source of information when we make casual observations follows from our everyday experiences. What we see and hear as we enact our various life roles provide the data of casual observation. We learn about people and events from participation in social life, and from gossip, old wives' tales, popular sayings and folk tales that have been passed along in our culture without formal codification. Such everyday occurrences as seeing a movie, a street fight, or a political event provide the new data of experience. Some observations are impressive and compelling but others are ignored because the source appears invalid. At any rate, information gleaned from casual observation is often the beginning of scientific hypotheses about people and their behaviour. For instance, we may know students who are 'achievers' in their college work and wonder what accounts for their excellence. It is often heard that they are talented, or intelligent, yet we may also observe that they seem different in other ways: even casual observation suggests that they spend a good deal of time at their studies; they appear to have well-organized study and work habits; some seem to hold a general attitude that achieving in school and in life is important. Some of these students appear to do well in almost anything they encounter. Others, however, may seem unable to achieve a social sense, the 'book-worms', those who spend a good deal of the time by themselves and are infrequently seen in the popular 'hangouts'. All these casual observations are a source of ideas for hypotheses about the factors associated with high achievement.

II Focused observation

1. *Case study method* (naturalistic observation). Early in the development of a science ideas are often formulated by studying one individual, or by observing organisms in their natural environments. Casual observations may be supplemented by these methods. Retrospective accounts of a subject's activities and behaviours may be collected from people who have known him well, or information may be obtained directly from a subject by asking him (interview). Assuming we were interested in 'high achieving' students, we could ask room-mates about the subjects' study habits and the way in which the subjects approach their school work. The self-reports of 'high achievers' could also furnish information about their study habits, past successes, attitudes towards achievement, etc. After studying several successful students in this systematic way (peer reports, personal interviews) we may discover consistencies in their behaviour that lead to more refined hypotheses about high achievement.

Subjects whose lives are more easily controlled than college students (school children, prisoners, military personnel, hospital patients) are accessible to *time-interval* study. In such a study, subjects are selected for observation during specific time-intervals; for instance, each hour they are observed for a five-minute period, and their behaviour is recorded under predetermined categories such as social-approach responses, aggressive responses, or other responses that are believed to be important in posing a problem and in finding solutions. A quantitative estimate of behaviour occurrences may be obtained from time-interval observation. For example, during a five-minute time interval sampling of pre-school children, boys showed three times as many overtly-aggressive responses as girls.

The information obtained from case studies and naturalistic observations has limitations similar to information obtained from casual observation. However, the focus is sharpened on the variables believed to be important. The size of the sample is still limited, however, and caution is required in generalising from a few individuals to many. Also, the numerous factors that may affect the characteristics one wishes to understand are not systematically controlled; some idea of their differential importance may be gained, but only enough for statement of a refined hypothesis. Quantification is slightly increased over casual observation; therefore, a gain in precision is expected, i.e., the *number* of hours in study, the *intensity* of attitude towards achievement, etc. The primary usefulness of case studies is in refining hypotheses before they are submitted to more rigorous test.

2. *Correlational analysis.* The purpose of a correlational analysis is to determine the relationship between two variables. We want to find out what

happens to one when another changes. Do they increase together? Does one increase as the other decreases, or does it make any difference?

In introducing the term *variable*, we mean that whatever it is we observe is measurable, and that it can be conceptualised as a dimension, e.g. height, weight, typing speed (words per minute), etc. Thus, to carry out a correlational analysis, it is necessary to provide measurements for the variables under study. For example, let us assume that casual observation leads us to believe that 'smart' people are 'high achievers' in college. To investigate the relationship between the degree of 'smartness' and the degree of 'college achievement', we must *define* each variable. It is important to stress that the definitions of variables in scientific work must not be mere verbal expressions. Definitions must be stated in terms of the 'operations' or the actions of the scientist with reference to the events under scrutiny. Thus an operational definition for 'weight' is the movement of a pointer on a standard balance in units of pounds and ounces. We may decide upon a standard IQ measure as the operational definition for 'smartness', and perhaps grade-point average as the operational definition of achievement. Thus, two measurements (scores) apply to our subjects. To complete the analysis it is necessary to obtain scores for our subjects on both variables. School records are a source of academic achievement scores (grade point), and administering an IQ test can provide a measure of intelligence. The remainder of the analysis entails statistical computations between the two sets of scores; the statistical results indicate the degree of relationship between achievement and intelligence as operationally defined.

Two statistics are commonly used to indicate relationship in psychological studies: the rank-order coefficient, rho (ρ), and the Pearson product-moment coefficient (r). Rho indicates how closely matched the two sets of scores are when both are ranked from high to low for all subjects. For example, if we had measures of achievement and intelligence on thirty students, we would rank our sample from highest to lowest on both of the variables. If our hypothesis was that higher intelligence indicates higher achievement, we would expect those students who rank high on the intelligence test also to rank high in school achievement. If this turned out to be the case, we would find a *positive* relationship between the two variables; that is, as one increases, the other increases, or, as one decreases, the other also decreases–they go in the same direction. However, if those who scored high on the intelligence measure scored low on the achievement measure, we would find a *negative* relationship; as one variable increases, the other decreases, and vice-versa. If there is *no relationship* between the two variables, the rank order coefficient would equal zero (or be very small); that is, when one variable changes it makes little or no difference in the direction of change in the other.

The Pearson product-moment coefficient is similar in theory to the rank order coefficient although the scores are not arranged into ranks; the pairs of raw scores are used by themselves and a different computational method is applied. The same explanation applies to the direction of correlation (positive, negative, or none).

The range of both coefficients is from plus one (1.00) to minus one (–1.00); plus 1.00 is a perfect, positive correlation where the highest member in IQ would also be the highest in achievement, second in IQ would be second in achievement, and so on. A minus 1.00 coefficient indicates just the opposite; number one in IQ would be last in achievement, number two in IQ would be next-to-last in achievement, and so on. Varying degrees of relationship lie in between these extremes. Statistical tables are available to help determine the degree to which chance factors may have accounted for the magnitude of the obtained coefficient.[1]

Information from correlational analysis has limitations. The quality is improved over the previous methods because the sample size is usually larger; thus, generalizations to a larger population are more appropriate. However, other variables that may influence the results have not been systematically controlled, and their effect cannot be determined. Correlational analysis *never* shows that one variable *caused* the change in the other. A third variable may cause the change in both variables. For example, if we correlated the number of umbrellas on the street with the number of puddles in the street, we are likely to find a high, positive correlation; however, it would be absurd to say that one caused the other; clearly, a third variable, the amount of rain, was antecedent to both observations. On the surface, it might appear that cause-and-effect is demonstrated through correlational analysis; however, it is quite hazardous to make the assumption that change in one variable causes the change in the other.

Although there are limitations to correlational analysis, it is still a potent investigative tool and may serve as a source of further hypothesis formulation. Because of the difficulties in submitting many human behaviours to rigorous control, a correlational analysis is sometimes the only feasible procedure.

III Controlled observation

Ideally, the more we are able to control variables, that is, keep them from interfering with our observations, the more likely we are to discover the conditions responsible for the changes we observe. The basic premise of controlled observation is: if all factors are held constant, except those we wish to observe, then any change observed is a result of the change in those factors; consequently, they must have caused the observed changes. The logic behind this premise is basically sound; however, in practice it is often extremely difficult to apply. The main problem lies in controlling (keeping constant) all of the variables that are not of immediate interest but which might affect the outcome of the experiment. Investigating human problems presents even added difficulty because ethical restrictions limit experimental control over a number of variables related to human behaviour, i.e. how the subjects are raised, their diet, their freedom of movement, etc., and all may affect the results of an experiment. For this reason, much controlled experimentation is carried out in a laboratory setting with animal subjects; however, psychologists more and more are developing suitable techniques of laboratory experimentation with humans. Controlled observations may be carried out in natural settings although the difficulties in controlling variables are multiplied because of the freedom of the subjects, and the inability to control the environment.

---

[1] For more detailed information on computing correlation coefficients and determining their probability of occurrence, the student should consult an introductory text in psychological or educational statistics.

To minimize the problem of controlling extraneous variables, several procedures are available. One is the method of *random selection.* Experiments are made on a *sample* of subjects, i.e. a sample of a population to which the experimenter wishes to generalize his results. If one is interested in the achievement of college students, the population may be defined as *all* college students; however, experimentation is done with only a few in hopes that the results will apply to all. If we select a sample of students *at random*, we maximize the chances that factors unknown to us, but which may influence the experimental results, will balance themselves. That is, from a sample selected randomly, the chances are good that no one characteristic will be favoured. For instance, in selecting students to investigate learning ability we would expect that a truly random sample would provide us with a range of intelligence, and random placement of subjects into two groups that receive different experimental treatments should cancel out the biasing effects of intelligence level. Each group should contain essentially the same overall level of intelligence. However, the validity of the assumptions behind random selection depends on having the entire population available for selection, an assumption that is rarely met in practice. Any restrictions on population availability weakens the chances that a truly random sample will be obtained.

Another method of overcoming difficulties in control is to establish *matched samples.* If an investigator knows some of the important variables affecting the events he wishes to study, he may equalize these variables in the groups by matching the subjects on them. For instance, if we are interested in the effect of monetary reward on the 'need for achievement' it is possible that the variables of intelligence and age are important; therefore we construct our groups so they are equal on measures of intelligence and on age. By actively equalizing the effects of such extraneous variables, we reduce the chance that they will have an effect on the experimental outcome.

Scientists further rely heavily on repeated *confirmation* to increase the certainty of their observations. If the same experimental findings are repeated (replicated) in a different setting, with different subjects, the likelihood is increased that the results followed from the manipulating of the antecedent variables. When findings in different settings are inconsistent, the experimental design must be examined for important variables that are not being controlled.

A *control group* forms an important part of a controlled experiment. The function of a control group is to provide an observation that cannot be attributed to the variable being manipulated. In examining the effect of a new teaching method on learning, for instance, it is necessary to observe two similar groups of children–one group being taught by the new method, and the other group, the control group, being taught in the usual way. At the termination of the experiment, the amount of learning shown by the group receiving the new method (the experimental group) is compared to the learning of the control group. Such a comparison furnishes assurance that the new method resulted in an increase over and above the increase that would normally occur over the same time interval.

To understand the controlled experiment further it will be helpful to look at an example. Suppose we are interested in designing an experiment to investigate high achievement, and we suspect that 'incentive' plays an important part in learning. Our hypothesis might be: if incentive is increased, then learning is increased. Thus, we are interested in observing the change in learning as we vary the amount (degree) of incentive. Since intelligence is likely to influence learning, we administer an IQ test to all subjects before the experiment and compose our experimental group and control group so both contain subjects of similar intelligence level. The experimental group is told that their performance on the task (learning a passage of material) will count towards their course grade. The control group is asked to learn the task, but not told that it will count towards their grade. After the subjects have studied the material we administer a test to observe how much they have retained. We then compare the amount of learning of the experimental group to the amount of learning of the control group. If the amount of learning of the experimental group is greater, our hypothesis is supported.

The variables in an experiment are referred to by specific names. In the example above, the *independent variable* is the degree of incentive. It is the variable we manipulate; we cause it to change in order to observe its effect on learning. In this study it is defined by the difference in instructions to the two groups, and we examined only two *levels,* i.e. *no* added incentive (control group), and *some* added incentive (experimental group). The degree of learning is the *dependent variable.* We wish to observe changes in this variable when experimental manipulations are applied. Changes in this variable are 'dependent' on changes of the independent variable. In the above case it is defined as the score on a test. Our other variable, intelligence level, is called a *controlled variable.* We equated it between the two groups in order to eliminate its effect on the dependent variable; its definition is the score on an IQ test.

The foregoing experiment was an example of the simplest controlled experiment, the *two groups design.* As with correlational analysis, specific statistical procedures are applied to the data in order to evaluate the possibility that the difference between the two groups could have occurred by chance. We test the difference between the mean scores (averages) of the two groups on the dependent variable to determine if they are actually different, or if their difference is such that it could be attributed to a chance occurrence. The computation of two statistics is generally used for the above purpose, the 'z' statistic, and the 't' statistic. If the sample size is large, greater than 30, the z-test is applicable; with smaller samples, we must use the t-test. Most elementary statistics books contain computational procedures for these statistics.

Controlled experiments may become more complex than the two groups design. A common procedure is to include more experimental groups so that multiple levels of the independent variable are examined. For instance, the above experiment could be expanded by including a third group that is told their achievement scores will apply to their grade and also that anyone obtaining a score over 80 per cent will be rewarded with $5.00. Thus, three levels of the incentive variable are investigated: (1) the control group receiving no extra incentive, (2) one group receiving a specific incentive, and (3) a third group receiving two forms of incentive. As a result, more is learned about the effects of incentive learning, and more precise statements about the findings can be formulated. If, for example, the three groups showed increased learning with increasing levels of incentive, the initial experimental hypothesis would be supported even further, i.e. increased incentive increases learning. However, assume that the two experimental groups did not differ in learning, although both were higher than the control group. In this case, the results would be qualified by saying that added incentive increases learning to a limited extent, but a point is reached where further incentive does not add to learning. A design such as this where one variable is examined on several levels, is called a *simple analysis of variance.*

Controlled experiments become more complex as more independent variables are introduced. To continue with the example, it is possible to examine the effect of incentive *and* 'achievement motivation' on learning. Achievement motivation can be measured by tests designed for that purpose. The subjects can be divided into groups of 'high achievers' and 'low achievers' based on their achievement motivation scores. Consequently, our design contains six groups of subjects matched on IQ, and varied on two independent variables, (*a*) achievement motivation, and (*b*) incentive. A group of subjects scoring high on achievement motive tests and a group having low scores would be exposed to each of the levels of incentive. The results would show the effects of achievement motive and incentive on learning, plus the 'interactional effects' between the two variables. For instance, an interactional finding might indicate that subjects who have high achievement motivation perform high in learning regardless of incentive level, while subjects with low achievement motivation show increasing performance as incentive increases. One would have evidence to conclude that incentive is useful for students who have not developed a personal need to achieve.

Most of the limitations of the controlled experiment have been discussed in relationship to the difficulty in obtaining control. In general, experiments that are performed in laboratory settings maintain better control than studies in natural environments. While the techniques of randomization and matched sampling help to reduce control problems in subject-variables, the natural environment presents more difficulties because of the numerous variables related to the setting. The more frequently experimental studies are confirmed, the greater the degree of certainty that the results are valid.

Others' observations

The observations of other investigators fall into the same categories that were discussed under personal observations. The way that others make their observations will reflect the same limitation, and uses, as they would if the student had made the same observations. One of the primary tasks in using others' observations is to evaluate the way in which they were made in order to determine their usefulness.

Literature is the primary source for locating others' results. The major handbooks, manuals and periodicals relating to psychology will direct the student to many of the major sources of information.

STATE INSTITUTE OF EDUCATION
Govt. of West Bengal

2 Reading research articles

T. R. Sarbin and W. C. Coe

In writing his term papers, the student psychologist will benefit from using both original sources and secondary sources. In general, original source material is to be found in the periodical literature, i.e. journals, reviews, bulletins and research monographs. Textbooks, handbooks, and encyclopedias give secondary source material in bringing together the content of the periodical literature and research monographs.

Research reports may appear forbidding to the beginning student. Technical language is freqently used, charts and graphs may be confusing, the concise form of reporting seems cold, and mathematical equations may be intimidating.

The purpose of this chapter is to make research papers less forbidding. This brief statement is intended to facilitate the reading of research articles so the student can become more involved in the elementary steps of the scientific enterprise. By clarifying the form of the usual article and by suggesting certain aids, these hints will give the beginning student some direction, some hope, and some skill.

Research articles generally follow a more or less standard pattern. In the first few paragraphs, *the introduction,* the author usually poses the problem to be investigated. His introductory statements should outline the problem in a general way. Frequently, he places it in a historical context by reviewing the propositions of previous investigations, or citing and quoting results and theoretical statements that are pertinent to his problem. Having given the reader a general perspective, the author becomes more specific and presents the particular aspects of the problem in which he is interested. At this point, the central feature of the investigation is expressed. It may be in the form of a narrative statement that certain observations will be related; it may be stated as a formal hypothesis; or it may be expressed as an informal guess or hunch.

The second section of a research article is concerned with procedures and is usually labelled *method.* The author describes in elaborate detail his apparatus, techniques or tests, characteristics of the experimental setting, his selection and description of subjects, his instructions to the subjects, and the experimental design. Detail is necessary in the methods section so that other investigators may confirm the findings (or question them) by repeating the experiment without altering the conditions.

Source: SARBIN, T. R. and COE, W. C. (1968) *The Student Psychologist's Handbook: A Guide to Sources.* Cambridge, Massachusetts. Schenkman, 47-51.

The authors and editors of psychological research reports have evolved a set of conventional symbols to designate certain standard features. The more commonly used symbols and their referents are:

N = the sample, the number of subjects ($N = 10$ should be read 'the number of subjects was ten')
S = subject
E = experimenter
O = observer

The next section is labelled *results* or *findings.* Here the data are presented, sometimes in expository form, sometimes in the form of tables and graphs, and sometimes by both methods. Because experiments usually use a large number of subjects, the general practice is to provide descriptive statistics of the data. When inferences are to be drawn from samples to populations, various statistical tests are included in order to evaluate the significance of the findings. In reading for acquaintance or for comprehension, the beginning student need not have a thorough grasp of descriptive or inferential statistics.

A *discussion* may follow the presentation of the findings. The author tries to link his observations with the original hypothesis, or guesses, and may speculate about implications for theory or applications to practice. The results of the study are compared with the findings of other investigators; where contradictions occur, the author will try to explain them. At the end of the article, the author customarily writes a condensed version of his research, a *summary.*

A research article may be read with any one of three aims: (1) acquaintance, (2) comprehension, or (3) critical evaluation. After preliminary selection based on the title or abstract, the student should approach each article with acquaintance as his first objective, comprehension as his second, and, if he has sufficient background and interest, critical evaluation as his third objective.

If the aim of the reading is *acquaintance* with a research paper, the student should first read the introductory paragraphs to determine the nature of the problem being investigated. Then the method and discussion should be scanned briefly. Finally, the summary statements should be carefully scrutinized. (In many journals, a summary statement appears in the form of an abstract at the beginning of the article.) By following this procedure, the student will achieve at least a cursory acquaintance with the contents of the article and perhaps acquire a 'feel' for what it is all about. The beginning student should be cautioned against becoming over-involved in the fine details of research presentation, lest he lose sight of the general conclusions and their implications.

With acquaintance as a basis, the student can now concentrate on acquiring a more profound *comprehension* of the work. In order to do this, he must scrutinize rather than scan the article. That is to say, he will ask a number of questions in the course of his reading. Some of the typical questions students have found useful in scrutinizing research articles are outlined in the following paragraphs.

The *introduction* will generally provide information to answer the following questions: What is the author trying to do? Is he trying to discover new relations among events? Is his work aimed at testing a hypothesis (a hunch) by systematic observation? Is he trying to confirm findings reported at an earlier time? Is he interested in determining the generality of a principle? Is he questioning a popular conception that has not been subjected to systematic examination through controlled experiments or careful observations? Is he interested in determining the fruitfulness of a method?

The *methods* section will generally provide answers to questions of this sort: How does the author convert his hunches or his beliefs into testable hypotheses? What operations does the experimenter employ to answer the questions to settle his doubts? How does the author specify the variables? For example, how does he justify his selection of an ink blot test as a measure of sexual interest? How does he select his subjects–are they randomly selected or selected according to some announced criteria? What kinds of manipulations or examinations of the subjects does he introduce? How are the control samples selected? Are the control samples adequate?

The *results* section will provide answers to these typical questions: What are the raw data from which the results are refined? Are they ratings, scale scores, speech samples self-reports, physiological measurements, test scores, census data, and so on? Have the data been processed to give summary statistics, such as averages? Or are the original protocols provided, such as electronic tracings, verbal responses to questions, etc.?

The *discussion* section will usually provide answers to questions of this type: Are the results in line with predictions from the author's initial hypotheses? How does the author relate his findings to conclusions reached by other investigators? How does he explain differences between his data and the predictions generated by his hypotheses? How does he integrate his findings with other current attempts to answer the same or similar questions?

Having achieved *comprehension*, the student is in a position to evaluate critically the research report by asking a few additional questions. Such questions cannot be asked by the beginning student; they are based in part on familiarity with research design, statistical methods, and the logic of experimentation. However, students at any level of sophistication may use common sense as the basis for criticism.

The student will ask the following questions to help him achieve a *critical evaluation*: Are the investigator's research hypotheses clearly linked to the questions raised? Are the specifications of the variables reasonable? Is he justified in specifying, for example, that a score on a personality test is equivalent to 'ego-control'? Is he concerned with representativeness both in selecting his subjects and in specifying the behaviours to be observed? In what ways are his controls adequate, and in what ways inadequate? Are the statistical tests appropriate to the data? May conclusions other than those of the author be derived from the same data? In what way are the findings significant, i.e. do they contribute in any important way to our understanding of conduct? Can the efficiency of prediction and control of behaviour be increased as a result of the application of the findings?

It is not unusual for the undergraduate student to hesitate before critically evaluating a research article. This hesitancy is understandable; for years he has practiced accepting the written word as the final authority. The reluctance to criticize may be eliminated if the student will bear in mind that a research article, simply because it is in print, is not the last word, nor is it perfect. In keeping with one of the purposes of this book–to help the student solve the problems that arise in connection with his creative efforts–the student should recognize that any research study provides only tentative and imperfect answers to questions. It is not improper for the student to be sceptical, to doubt, and to raise questions. By asking penetrating questions, new hypotheses emerge and new research is undertaken. This is as it should be: new research is the hallmark of science.

To summarize: The student should first attain cursory knowledge of an article by scanning the introduction, reading the summary and conclusions, and superficially examining the rest. Reading for comprehension and for critical evaluation may follow. Some suggestive questions have been listed above as an aid to the student who seeks comprehensive understanding of research articles.

3 Nursery teachers' perceptions of their pupils: an exploratory study

B. Thompson

1. Introduction

At the present time strong pressure is being brought to bear upon the government to increase the provision of pre-school education. While this seems right in principle, it must be admitted that very little hard evidence exists as to what goes on in nursery schools and the subsequent effects on the children.

In general, British nursery schools provide children with an enriched environment with which they can interact freely under the eyes of sympathetic adults. Their aims, as described by Taylor *et al.* (1972), are the socio-emotional, intellectual, aesthetic and physical development of the child, and the creation of an effective transition from home to school. However, the question of how the teachers behave in order to achieve these aims is as yet largely unknown. A few studies, notably that of Gardner and Cass (1965) and the more recent work of Marianne Parry and Hilda Archer (1974), have attempted to describe the role of the nursery-school teacher, but because they have concentrated on 'best practices' they have tended to present a somewhat rosier picture than is in fact typical.

It would be generally agreed that the terms in which teachers see their pupils strongly influence the ways in which they behave towards them, but no attempt has yet been made to explore this area–although Nash (1973) has studied the main constructs used by teachers in an examination of the relation between teachers' attitudes and pupils' performance. The repertory grid technique developed by Kelly (1955) offers a method of exploring the ways in which a person categorizes others and, therefore, the expectations he has of them. The work described in this paper, which uses the grid in an exploratory study of the ways in which nursery teachers perceive their pupils, is an initial attempt to chart a relatively unknown area.

2. The context of the investigation

Despite the fact that most nursery heads hold to a common set of

---

Source: First published in this volume. This investigation was carried out while the author was a member of the NFER (National Foundation for Educational Research in England and Wales) Pre-school Project research team, which was engaged in introducing a structured language programme into nursery schools and evaluating its effect on children's performance.

principles, the schools they run often differ considerably. Presumably such differences affect how the teachers see their pupils and act towards them. This study was carried out in two nursery schools within the same area, which gave very different impressions to the observer.

School 1 had eighty pupils on its roll, ten of whom attended full time. The parents of the majority of the children were in skilled or semi-skilled occupations, and roughly a fifth were of West Indian, Indian or Pakistani origin. The permanent staff consisted of one qualified teacher, who was the headmistress, and two very experienced nursery assistants. The school was at the time participating in a research project concerned with evaluating the effects of a structured language programme on the linguistic and perceptual skills of pre-school children. There was a very warm, permissive atmosphere in the school and, except when attending their twenty-minute language sessions, children were largely left to choose their activities.

School 2 was a larger nursery with 120 pupils on its roll, approximately twenty of whom were full time. The social backgrounds and ethnic origins of the children were very similar to those of the School 1 children, although a rather larger proportion were not English-speaking. The staff consisted of the headmistress (who did not participate in the study, but gave it her blessing), a qualified teacher and three nursery assistants. The head was unhappy at the idea of structured teaching at the pre-school level and was carrying out her own language stimulation programme in a more traditional manner. She was also interested in the development of scientific concepts and provided a lot of equipment to enable the children to explore these. The school hummed with purposeful activity.

3. The method used

a. Rationale

It would have been possible to explore the terms in which teachers saw their pupils by means of traditional attitude scales, to be completed by all staff members. However, any scale of this sort has implicit in it the assumptions of the compiler as to which characteristics of the children are important. The teachers would probably have been able to state their position on every item included in such a scale, but it is likely that for each person some of the items would have had very little relevance and other important dimensions would have been omitted. These problems could have been partially solved by obtaining a rating of salience against each item and allowing teachers to add items which they considered should have been included, although taking all this into account in the subsequent analysis would have been quite difficult.

In order to counter these objections, it was decided to use a method which elicited from each staff member the terms in which she saw her pupils, but to add to those a limited number of attributes which were considered of importance by the members of the research team.

b. The Kelly Repertory Grid

The Kelly Repertory Grid (Kelly, 1955) is an instrument designed to elicit from individuals which characteristics (dimensions) they use to categorize a certain aspect of their environment, and how these dimensions relate one to another. That is to say, it attempts to understand from each individual the particular terms in which he sees his world.

The grid is completed as follows:[1]

(i) The subject is asked to name a certain number of items in the area under consideration (i.e. in this study, a number of nursery-school pupils).
(ii) These are entered on numbered cards.
(iii) Three cards (a triad) are then shown and the subject is asked to think of some important way in which two of the items are like one another and differ from the third.
(iv) The descriptions of the pair and of the singleton are entered to the left of the grid (see Figure 1, p. 20).
(v) The cards are then sorted into those representing items similar to the pair and those similar to the singleton. The result of the sorting is then entered on the grid, with ticks indicating those like the pair and crosses indicating those like the singleton.
(vi) The procedure is repeated with new triads.

The items on the cards to be sorted are referred to as elements and the categories into which they are sorted are called *constructs.*

The completed grid can then be analysed to show relations between constructs or between elements.

Although more difficult and time-consuming to complete than the traditional form of attitude scale, the repertory grid is frequently found to be more acceptable to respondents. They appear gratified that someone is seriously concerned to know what they think. They are also forced to think in depth about a particular aspect of their environment, and this proves to be a challenging and interesting task.

c. Methods of administration used in this study

The procedure described in section (*b*) above was followed, with some slight variations for the two schools.

---

[1] A number of variations in procedure are possible: see Bannister and Mair (1968) Chapter 2, pp. 59–67.

	Constructs		Elements (pupils)																							
	Two are the same	One is different	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
			F	M	M	M	F	M	F	M	M	F	M	F	F	F	M	M	F	F	M	F	F	M	F	M
A	open personality	quiet	✓	✓	×	×	✓	×	×	✓	×	×	×	✓	✓	✓	✓	✓	✓	×	✓	×	✓	✓	×	✓
B	quiet but thoughtful	chatters without thinking	✓	✓	✓	✓	×	✓	✓	✓	✓	✓	✓	✓	×	✓	?	×	×	×	×	✓	×	×	✓	×
C	easily embarrassed	less easily embarrassed	×	×	×	×	×	✓	✓	✓	×	×	×	×	×	×	×	×	×	✓	×	×	×	×	×	×
D	more mature	less mature	✓	✓	✓	✓	×	×	✓	✓	✓	✓	×	✓	×	✓	×	×	×	×	✓	✓	✓	×	✓	✓
E	unreliable	reliable	×	×	×	×	✓	×	✓	×	×	×	×	×	✓	×	✓	×	✓	×	×	×	×	✓	×	×
F	uninhibited	inhibited	✓	✓	×	×	✓	×	×	✓	×	×	×	✓	✓	✓	✓	✓	✓	×	✓	×	✓	✓	×	✓
G	friendly	shy	✓	✓	×	×	✓	×	×	✓	×	×	×	✓	✓	✓	✓	✓	✓	×	✓	×	✓	✓	×	✓
H	good concentration	lacks concentration	✓	×	✓	✓	×	×	✓	✓	✓	✓	✓	✓	×	✓	×	✓	×	×	✓	✓	✓	×	✓	✓
I	good conversation	poor conversation	✓	✓	×	×	×	×	✓	✓	×	×	×	✓	✓	✓	×	×	✓	×	✓	×	✓	×	×	✓
J	good activity	little activity	✓	✓	✓	×	✓	×	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	×	✓	✓	✓	✓	✓	✓
K	constructive	poor constructive ability	✓	×	×	✓	?	×	×	✓	×	✓	×	✓	✓	✓	×	×	✓	×	✓	✓	×	×	✓	×
L	imaginative play	less imaginative play	✓	×	✓	×	✓	×	✓	✓	×	×	×	✓	✓	×	✓	✓	✓	×	×	×	✓	✓	×	✓
M																										
N																										
O																										
P																										
Q	likely to succeed at school	less likely to do so	✓	✓	✓	✓	×	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	×	✓	✓	✓	×	✓	✓
R	frequently talks to other children	seldom does so	✓	✓	×	×	✓	×	✓	✓	×	✓	×	✓	✓	✓	✓	×	✓	×	✓	×	✓	✓	×	✓
S	independent	dependent	✓	✓	✓	✓	✓	✓	×	✓	✓	✓	×	✓	✓	×	✓	✓	✓	×	×	✓	✓	✓	✓	✓
T	spends a lot of time with puzzles, table materials etc.	spends little time with them	✓	×	×	✓	×	×	✓	✓	×	✓	×	✓	×	✓	×	×	✓	×	×	✓	✓	×	✓	×

Fig. 1 Example of a repertory grid

(i) *Selection of elements*
In School 1 the headmistress was asked to produce a list of twenty-four pupils whom she felt would be well known to all three members of staff. She was asked to choose roughly equal numbers of boys and girls and to include as wide a variety of children as possible, not only those who were outstanding in some way. Three sets of 24 cards were then made up, each bearing the name of the child and a number from 1 to 24.

In School 2 it was not possible for all four members of staff to respond to the same group of children. As this was a larger nursery, they were less likely to be well enough acquainted with children outside their own rooms. Therefore each staff member sent in her own list of children, and sets of cards were made out accordingly.

(ii) *Completion of grids*
In School 1 the grid was administered to all three members of staff together. As periods when they could all be free together were of rather short duration, several sessions were required.

Eight triads were presented, every child being included once (constructs A-H, Figure 1).

The teachers were then asked if there were any characteristics they used in thinking about their pupils which had not so far been elicited. If such characteristics existed they were asked to sort the children according to these and enter them in the usual way.[3]

Finally, they were told that there were four characteristics in which the research team was particularly interested and asked to sort the children according to these.

The four introduced constructs were:

Q likely to succeed at school/less likely to do so
R frequently talks to other children/seldom does so
S independent/dependent
T spends a lot of time with puzzles, table materials, etc./spends little time with them.

If they had already used any of these, they were to write in 'as A', 'as B', etc., (the letter of the relevant construct).

Group administration was not possible in School 2 and, in view of previous experience, was not thought to be particularly advisable. The method was amended as follows.

Two teachers were seen together and the other two on their own. The sex of the children was entered on the grid, the first three triads presented

---

[3] At this point the disadvantage of group administration became evident, in that there was a tendency to enter fewer constructs than they were capable of so as not to keep their colleagues waiting.

95499

and related sortings carried out. Each member of staff was then given a written sheet of instructions, grid and pack of cards, and asked to complete the grid in her own time. Explanations were given of the introduced constructs and the opportunity for staff to add constructs of their own in addition to those elicited by using the triads.

4. The results

a. Comparison of the verbal content of the individual grids

In this section, the data is examined in order to answer two broad questions. First, in what terms do these nursery teachers see their pupils, i.e. what are the characteristics and behaviours to which they attend; and second, to what extent do their perceptions differ from one another?

Ninety-four constructs were elicited in all, and a classification system was produced on the basis of these (Table 1). Any such system is, of course liable to be in some way idiosyncratic to the research worker. Hence later analyses look at the statistical relations between the constructs of the various subjects.

Table 1 *Classification of elicited constructs*

A		*Constructs concerned with personal qualities*
	1	introversion/extraversion
	2	confidence, maturity
	3	sense of humour
B		*Constructs concerned with social behaviour*
	4	towards peers
	5	towards adults (discipline)
C		*Constructs concerned with competence and ability*
	6	concentration, attentiveness
	7	observant, constructive, easily grasps problems, etc.
	8	conversation and speech
	9	specific nursery activities
D		*Miscellaneous constructs*
	10	physical activity
	11	background and origin

No single teacher produced constructs in all the listed categories, and quite marked differences were found between teachers in the two different schools. A detailed description is first given of the construct system supplied by one teacher in School 1, which can then act as a basis for comparison with the others.

This teacher sees children primarily in terms of their personal qualities, seven out of her twelve elicited constructs falling into categories A1 and A2. Thus she sees her pupils as:

open personality/quiet
chatters without thinking/quiet but thoughtful
less easily embarrassed/easily embarrassed
uninhibited/inhibited
friendly/shy

All these constructs can be seen as describing a dimension similar to that referred to by psychologists as extraversion/introversion. She also sees them as:

more mature/less mature
reliable/unreliable

i.e. constructs concerned with maturity.

Four other constructs fall into the competence and ability categories:

good concentration/lacks concentration
constructive/poor constructive ability
good conversation/poor conversation
imaginative play/less imaginative play

Her colleagues in School 1 produce fairly similar sets of constructs, as shown in Table 2. It is important to note that all nursery activities referred to by teachers in this school concerned constructional activities and creative and imaginative play.

Table 2 *Percentage of constructs supplied by teachers in School 1 falling into each category*

Categories		*Teachers* 1A	1B	1C	All
A1	extraversion/introversion	42	30	36	36
A2	confidence, maturity	17	30	36	28
A3	sense of humour	–	–	–	–
		49	60	72	64
B4	social–peers	–	–	–	–
B5	social–adults	–	–	–	–
		0	0	0	0
C6	concentration, etc.	8	10	–	6
C7	observant, etc.	8	–	–	3
C8	conversation	8	10	9	9
C9	nursery activities	8	10	18	12
		32	30	27	30

19.7.93
7079

D10	physical activity	8	10	–	6
D11	background, origin	–	–	–	–
		8	10	–	6

The four teachers in School 2 seem to stress personality qualities (Category A) rather less than those in School 1. They do, however, produce a number of constructs concerned with social behaviour, both in peer and adult/child interaction. For example, Teacher A in this school gives:

friendly to adults/tendency to withdraw from adults
excitable/more placid
insecure in a strange situation/responds well to change

as well as:

sociable, usually in a group/tends to play alone or with one friend
a leader/a follower
always joins in group activities/joins only occasionally or prefers to watch

and:

aggressive tendency/lacking in aggression[3]
rather unhelpful/likes to be helpful
resistant to discipline/cooperative

In addition to these she produced five competence/ability constructs:

lacking in concentration/concentrates quite well
less observant/quite aware of surroundings
rather slow to understand instructions/grasps new ideas or instructions rapidly
language problems/uses words quite well
enjoys books/spends little time with books

and finally, two concerned with background and origin:

slight behaviour problems–poor background/normal child from a reasonable home
immigrant/non-immigrant

Three things strike one about this system as compared with the one described above:

(*a*) there is much more negative evaluation;
(*b*) there is more stress on cognition;
(*c*) there is a concern with background.

---

[3] Constructs such as these show the inevitable dilemma of such a categorization system. The first of these constructs could clearly be classified in category B5, whereas there is no indication as to whether the second should be classified under B4 or B5.

There is less similarity between the constructs produced by the members of staff in School 2 than was the case for School 1. However, the tendency for greater negative evaluation–as expressed in constructs concerned with lack of discipline–is present in all of them, and three out of four produce one or more constructs concerned with background and origin.

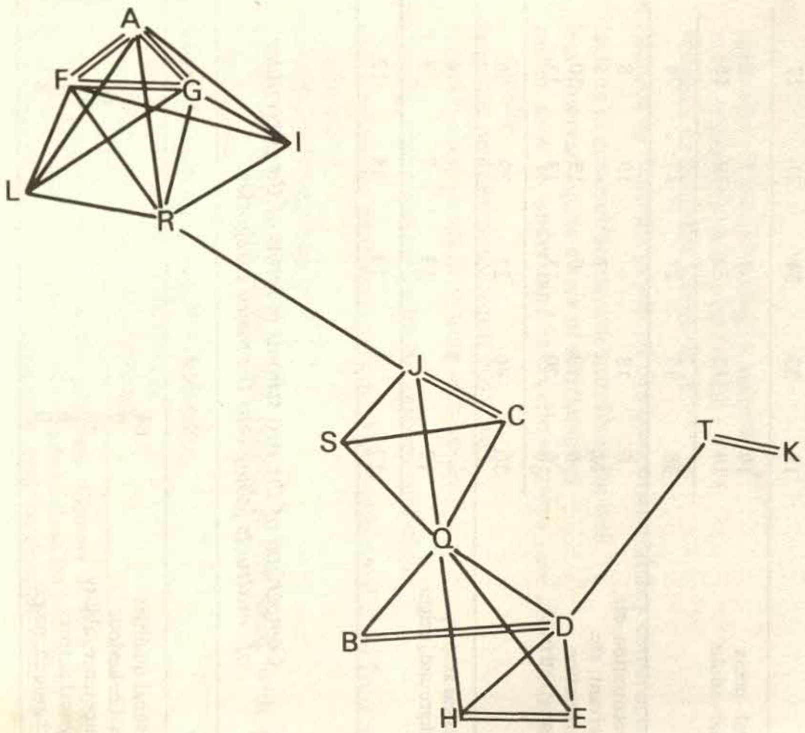
Table 3 shows the classification of constructs provided by teachers in School 2, and those of the two schools are compared in terms of major categories in Table 4.

Table 3 *Percentage of constructs supplied by teachers in School 2 falling into each category*

		Teachers				
Categories		2A	2B	2C	2D	All
A1	extraversion/introversion	12	20	7	7	12
A2	confidence, maturity	6	13	20	7	12
A3	sense of humour	–	–	7	7	3
		18	33	34	21	27
B4	social–peers	19	–	7	7	8
B5	social–adults	19	13	20	20	18
		38	13	27	27	26
C6	concentration, etc.	6	13	–	13	8
C7	observant, etc.	12	–	7	–	5
C8	conversation	6	13	7	13	10
C9	nursery activities	6	20	13	13	13
		30	46	27	39	36
D10	physical activity	–	7	–	7	4
D11	background, origin	12	–	13	7	8
		12	7	13	14	12

Table 4 *Comparison of the two schools in terms of the percentage of constructs falling into the major categories*

Categories		*School 1*	*School 2*
A	personal qualities	64	27
B	social behaviour	0	26
C	competence, ability	30	36
D	physical activity	6	4
	background, origin	0	8



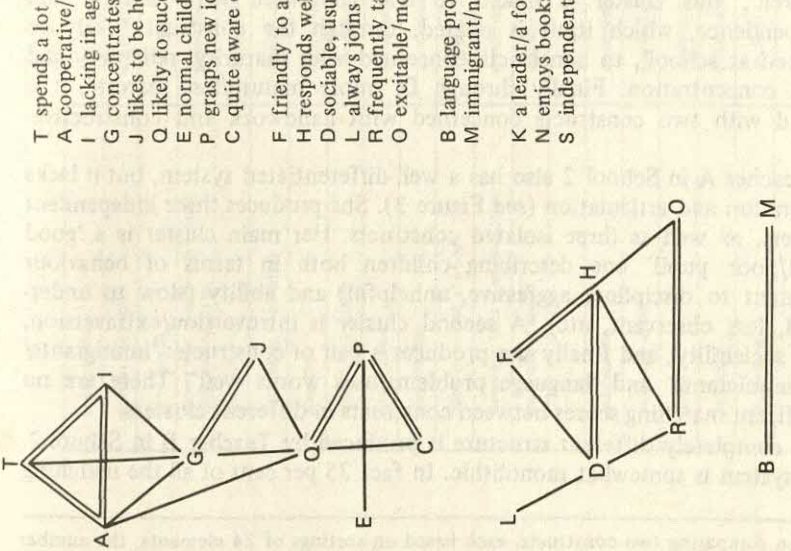
A open personality/quiet
F uninhibited/inhibited
G friendly/shy
I good conversation/poor conversation
L imaginative play/less imaginative play
R frequently talks to other children/seldom does so

J good activity/little activity
C less easily embarrassed/easily embarrassed
S independent/dependent
Q likely to succeed at school/less likely to do so

B quiet but thoughtful/chatters without thinking
D more mature/less mature
H good concentration/lacks concentration
E reliable/unreliable
T spends a lot of time with puzzles, table materials, etc./spends little time with them
K constructive/poor constructive ability

Fig. 2 Diagram showing significant relations* between constructs–Teacher 1A

*A double line indicates that this is the highest relationship for both constructs.



T spends a lot of time with puzzles, table materials, etc./spends little time with them
A cooperative/resistant to discipline
I lacking in aggression/aggressive tendencies
G concentrates quite well/lacking in concentration
J likes to be helpful/rather unhelpful
Q likely to succeed at school/less likely to do so
E normal child from a reasonable home/slight behaviour problems–poor background
P grasps new ideas or instructions quickly/rather slow to understand
C quite aware of surroundings/less observant

F friendly to adults/tendency to withdraw from adults
H responds well to change/insecure in a strange situation
D sociable, usually in a group/tends to play alone or with one friend
L always joins in group activities/joins in only occasionally or prefers to watch
R frequently talks to other children/seldom does so
O excitable/more placid

B language problems/uses words well
M immigrant/non-immigrant

K a leader/a follower
N enjoys books/spends little time with books
S independent/dependent

} isolated constructs

Fig. 3 Diagram showing significant relations between constructs–Teacher 2A

b. The structure of the grids

As was stated above, a verbal content analysis always means that the classification system of the research worker plays a dominant role. In order to correct for this a simple statistical analysis was carried out. Matching scores were calculated between all pairs of constructs and, from the resulting matrices, cluster diagrams were drawn showing all those matching scores which reached a certain level of significance.[4] These diagrams give two basic types of information:

(*a*) they show how constructs relate to each other to form 'meaning structures';
(*b*) they show how well differentiated and articulated the systems are.

In this section, three of the systems are taken as illustrations of the above points. Teacher A, from School 1, produces a highly differentiated and yet well-integrated structure (see Figure 2). That is to say, she produces a number of distinct construct clusters which are linked one to another.

Her first cluster is clearly an extraversion/introversion one, but includes two constructs concerned with conversation. This would suggest that she considers these to be more indicative of personality than of competence and ability. Through one of these constructs, 'frequently talks to other children', this cluster is linked to one concerned with activity and independence, which itself is related, through the construct 'likely to succeed at school', to constructs concerned with maturity, reliability and good concentration. Finally through D, 'more mature/less mature', it is linked with two constructs concerned with handwork and constructiveness.

Teacher A in School 2 also has a well-differentiated system, but it lacks integration and articulation (see Figure 3). She produces three independent clusters, as well as three isolated constructs. Her main cluster is a 'good pupil/poor pupil' one describing children both in terms of behaviour (resistant to discipline, aggressive, unhelpful) and ability (slow to understand, less observant, etc). A second cluster is introversion/extraversion, with sociability, and finally she produces a pair of constructs: 'immigrants/non-immigrants' and 'language problems/uses words well'. There are no significant matching scores between constructs in different clusters.

A completely different structure is produced by Teacher B in School 2. Her system is somewhat monolithic. In fact 35 per cent of all the matching

---

[4] When comparing two constructs, each based on sortings of 24 elements, the number of similarly sorted elements is counted and 12 (i.e. the number of matches expected by chance) subtracted from the resulting figure. Matching scores of 6 or more, regardless of sign, are considered significant. This method is an extension of McQuitty's (1957) elementary linkage analysis.

			1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*	*19*
concentration	no concentration	*1*		10	9	7	7	6	9	8	7		6								
good at handwork	not so interested in it	*2*	10		9	7	7	6	9	8	7		6								
good speech	bad speech	*3*	9	9		6			8	7	6	6		6							
spends a lot of time with puzzles, table materials, etc.	spends little time with them	*4*	7	7	6																
responsible	irresponsible	*5*	7	7					6	7	6			6			6				
likes plenty of attention	doesn't command staff attention	*6*	6	6									6								
attentive	innattentive	*7*	9	9	8		6			11	8	8	6	8	7		6				
friendly and willing	withdrawn	*8*	8	8	7		7		11		7	7		7							
domesticated	not domesticated	*9*	7	7	6		6		8	7		10	7	6					6		
very talkative	no communication at all	*10*			6				8	7	10			8							
likely to succeed at school	less likely to do so	*11*	6	6				6	6		7				6						
frequently talks to other children	seldom does so	*12*			6		6		8	7	6	8				7	8	6			
more sure of herself	very shy	*13*							7				6							7	
musical	not so musical	*14*												7			7	7	7		
independent	dependent	*15*					6		6					8		7		8		6	
physically active	physically inactive	*16*												6		7	8		6	6	
spiteful	considerate of others	*17*									6					7		6		8	
boisterous	quiet and reserved	*18*													7		6	6	8		
disobedient	obedient	*19*																			

Table 5 *Matrix of significant matching scores produced by Teacher 2B*

scores reach a level which is statistically significant, as compared with 23 per cent for teacher 1A and 12.5 per cent for Teacher 2A. Because of the very high degree of interrelation it is difficult to produce a satisfactory diagram, so I have given instead a matrix with the constructs reordered to show the main groupings, and with the non-significant matching scores omitted (see Table 5). From this it can be seen that she produces a massive 'good pupil/poor pupil' cluster, consisting of constructs 1-12 which, through construct 12, links to a sub-cluster (constructs 13-18) distinguishing between quiet, considerate, dependent children and those who are boisterous, spiteful and independent. Finally, construct 19 has no significant links.

These three examples should make clear the varying amount of differentiation in people's construct systems even in an area central to them. They also show how inadequate verbal classification is as a way of representing the frames of reference for each individual teacher.

It is perhaps of interest to note in passing one major difference between these nursery teachers and a group of secondary teachers investigated by the writer. As can be seen from the above, nursery teachers seem to think that it will be the friendly, uninhibited extraverts who will succeed at school. In the great majority of cases, secondary-school teachers find introverts the better pupils, possibly because they are easier to manage in a more formal classroom situation.

c. *The introduced constructs*

It will be recalled that the research team introduced four additional constructs into the grids in order to explore further certain areas of particular concern.

(i) *Frequently talks to other children/seldom does so*

The researchers were also concerned with the introduction and evaluation of a structured language programme in nursery schools. They therefore wished to understand what part children's talking played in the construct systems of their teachers.

This construct is well integrated into the systems of every teacher in the sample, which suggests that they found it meaningful. What is more important to know, however, is the part that it plays in their construct systems. That is to say, whether teachers see children who talk a lot as competent and likely to be successful pupils. Of the forty significant matches with this construct overall, nineteen are with constructs in category A, i.e. those concerned with the personal qualities of the children. In fact, one of the reasons why this construct is so well integrated may be because of its relationship with the introversion/extraversion dimension. This is particularly the case for School 1, where ten out of its sixteen significant matching scores are with constructs indicating this dimension. The remaining six are related to: other speech abilities (three), activity level (two) and only one to a construct in the cognitive category not concerned with speech–'imaginative play'.

In School 2, although there are a number of significant links to constructs in category A, there are the same number related to the cognitive category, confirming the original impression that teachers in this school are more cognitively oriented.

When this study was carried out in School 1, the language project had scarcely got under way. It would be interesting to see if the staff still considers that conversation bears little relation to cognitive abilities, now that they are engaged in helping children improve their language, and through it their cognitive skills.

(ii) *Spends a lot of time with puzzles, table materials, etc./spends little time with them*

Handwork, puzzles and creative work generally play an important part in the activities of the nursery school. In the course of the language project the team was trying to increase the children's manual and perceptual skills by extending these activities. (This was not the case in School 2 which, it will be recalled, was not taking part in the project.) This construct was introduced in order to discover whether teachers saw performance on such tasks as related to cognitive ability, and hence whether improved performance would improve the children's chance of success at school.

This construct was less well integrated into the spontaneously produced construct systems. Three teachers related it significantly to only one other construct, three to three others, and only one integrated it into the major dimension of her system.

The great majority of the significant relations are with constructs which fall into category C, i.e. competence and ability. When one examines these in greater detail, however, one finds that in School 1 three out of the five significant relations are to another construct about handwork or creativity. Only Teacher B in this school sees handwork as related to wider abilities, e.g. 'good concentration'. In School 2 handwork seems to be seen as much more a part of school-related abilities. Three teachers see it as related to concentration and two to good speech, while for one it is related to another introduced construct, 'likely to succeed at school'.

This construct again reflects the difference between the two schools in terms of the importance they see in cognitive development.

(iii) *Independent/dependent*

Most nursery teachers stress that one of their main aims is to develop independence in their pupils. It is not, however, really clear what they mean by this, so this construct was introduced to try to elucidate the matter.

One teacher in School 2 does not see this construct as related to any other in her system, suggesting that it may be relatively meaningless for her. For a second teacher in the same school, it is only related to 'active/inert'.

Since two other teachers see a relation to activity, this is clearly not a random relation, but it does little to clarify its meaning. Further examination of related constructs suggests that for four subjects it is related to 'talks to other children', and to confidence and extraversion. The independent child seems to a large extent to be active, outgoing, confident and talkative, but only two teachers see this type of child as 'likely to succeed at school'.

(iv) *Likely to succeed at school/less likely to do so*
One of the functions of the language project was to attempt to improve the quality of the preparation children are given for school. It was clearly important to discover which qualities in the children the teachers saw as predictive of future success at school. If these were cognitive abilities, one might expect the teachers to be interested in the project; but if they believed that personal and social qualities were all-important, they would be less appreciative of the language programme. In the event it was disappointing that likelihood of success in school was less integrated into the systems of the teachers in School 1, where the language project was active, than in those of School 2. One of the reasons was the apparent unwillingness of the teachers in School 1 to classify pupils as less likely to succeed. Only three, four and five pupils, respectively, were so classified by these teachers– almost as though they were superstitiously afraid of blighting the children's future. Even in School 2, the figures were seven, seven, seven and eight, which are rather low. One might have expected a much more even split on a 'more/less' construct such as this, and it raises doubts about the willingness of the teachers to face reality.

Despite this, the construct does appear to have meaning for the majority of the teachers. For all four teachers in School 2, it is integrated into quite a major cluster, as is the case for Teacher A in School 1. The latter, however, sees it related almost entirely to personal qualities, the only exception being 'good concentration/lacks concentration'. The other two School 1 teachers are interesting in that one sees it as related to independence while the other sees it as related to children not having 'minds of their own'. In School 2, again, the more cognitive orientation is apparent. For these teachers the probability of success at school is related mainly to concentration, awareness of surroundings, attentiveness and, to a lesser extent, talkativeness and good speech. Only two of them see it as significantly related also to the discipline constructs they offered.

5. Summary and conclusions

This paper attempts a preliminary exploration of how nursery teachers see their pupils, using the Kelly Repertory Grid to elicit from each teacher the dimensions she considers to be of importance.

At a purely verbal level, the staff classify children broadly in terms of personal qualities, social behaviour, competence and abilities, physical activity and background.

Of the schools investigated, the teachers in School 1 can perhaps be said to have a more 'tender-minded' view of their pupils. They give no negatively evaluative constructs and find it difficult to see children as unlikely to succeed in school. Their major concern is with personal qualities such as extraversion/introversion, confidence, independence and maturity, qualities which are in line with the traditional philosophy of British nursery schools.

Teachers in School 2, on the other hand, have a much tougher attitude to their pupils. They appear to be concerned with control and discipline, and are quite prepared to see children as aggressive and spiteful. They also have a more cognitive approach, and see cognitive abilities as essential for success at school. Not only do the schools differ in their perception of their pupils, but so do the individual teachers, as can be seen by comparing their construct systems. To be of service to the individual, the construct system should be both differentiated and integrated. Of the seven teachers investigated here, two can be said to have only one major dimension along which they view pupils, 'good pupil/bad pupil'. Monolithic structures such as these are not very helpful guides for action, as their diagnostic procedure has two categories only. Two more teachers in the sample have fragmented systems. It is possible that these are an artefact of the technique, i.e. that if the elicitation procedure had gone on longer, some linking constructs might have been produced. The problem with systems of this kind is that the teacher may be faced with unresolvable conflicts when children fall at different poles in various clusters. Only if the system is differentiated *and* integrated is the teacher able to understand and treat her pupils consistently and yet according to their individual needs.

From this study we have some idea of what a group of nursery teachers expect of a good pupil. How far do their expectations coincide with those of the infant teachers for whom they are preparing their pupils? Further research into this area is essential if the quality of pre-school education is to be improved further.

References

BANNISTER, D. and MAIR, J. M. M. (1968) *The Evaluation of Personal Constructs.* New York: Academic Press.

GARDNER, D. E. M. and CASS, J. E. (1965) *The Role of the Teacher in the Infant and Nursery School.* Oxford: Pergamon Press.

KELLY, G. A. (1955) *The Psychology of Personal Constructs.* New York: Norton.

McQUITTY, L. L. (1957) 'Elementary linkage analysis for isolating orthogonal and oblique types and typal relevance, *Educational and Psychological Measurement,* Vol. 17.

NASH, R. (1973) *Classrooms Observed.* London: Routledge and Kegan Paul.

PARRY, M. and ARCHER, H. (1974) *Pre-school Education.* London: Schools Council Research Studies, Macmillan.

TAYLOR, P. H., EXON, G. and HOLLEY, B. (1972) *A Study of Nursery Education,* Schools Council Working Paper 41. London: Evans/ Methuen Educational.

Section Two
Personality theories and dimensions

Introduction

The scientific study of personality goes back no more than a hundred years, but a number of theories have been proposed in that time and many experiments carried out to support or refute them. Yet no one theory has been wholly supported and few have been wholly discarded. So in this section we cannot even attempt to provide a representative selection; instead, we offer a flavour of work in two particular areas of the field, together with some evaluation of each. The two positions selected, the psychoanalytic and the factor-analytic, represent respectively the idiographic and nomothetic approaches. The idiographic method normally proceeds by the examination of individual cases, out of which a theory is inductively built up; on the other hand, the nomothetic approach attempts to establish general rules from the study of large groups of individuals, and then proceeds to the systematic testing of the hypotheses which follow from these rules, to see if the predictions are supported. Both methods have their weaknesses as well as their strengths: the psychoanalyst may have a tendency to over-generalize from his small number of cases, while the factor-analyst may not be able to tell us much about any particular individual.

This section begins with five chapters from Freud's *An Outline of Psycho-analysis,* in which he introduces us to some of the 'inhabitants' of his world. They include the id, ego and super-ego, the instincts of love and destruction, the sexual stages, the qualities of consciousness, unconsciousness and the pre-conscious, and the dream and its interpretation. These are followed by two extracts from the writing of Paul Kline. The first takes one of Freud's ideas, that of the Oedipus complex, and considers the main studies which have attempted to examine its validity. In the second, Kline looks at an attempt to validate in Great Britain an American questionnaire designed to measure orality; the outcome of the study demonstrates some of the pitfalls in going for a quick refutation (or indeed confirmation) of Freudian ideas.

In the other three extracts we move to the factor-analysts. Cattell attempts to show how the use of his sixteen main personality factors, in conjunction with measured ability and level of motivation, might go almost the whole way to predicting an individual's school achievement. Here we have the attempt at rigorous scientific analysis, rooted in replicable, observable, quantified data. Cattell describes convincingly, too, how his work could be used sensitively in child-guidance and advisory situations.

In the following article Eysenck criticizes Cattell's theory of primary and second-order factors and argues that his own approach of using fewer factors to measure personality is more viable. If we accept his arguments, which are based on detailed data analysis, then we might agree that the two dimensions of extraversion and neuroticism could be enough for most purposes.

Finally, Entwistle reviews recent studies of the relation between personality dimensions and academic attainment, studies using scales developed by either Cattell or Eysenck. He highlights contradictions in the findings, discusses some of the possible causes of these and concludes by specifying those areas on which future research might most usefully concentrate. His paper demonstrates that the factor-analytic approach to the study of personality and its relation to school achievement still has a long way to go; we are still very uncertain how far typical classroom practice affects the performance, or even the personality pattern, of the individual–in terms, that is, of this type of measurement.

In fact, much of our work in the personality field, whatever school we lean towards, is still fairly theoretical. We are still far short of the kind of analyses or predictions that would enable us to intervene with any degree of certainty in the career of an individual. But there are some generalizations that we can begin to make with rather more confidence. So far, the psychoanalytic orientation has had most relevance, in the educational context, to our knowledge of what is wise or poor handling of individuals, particularly when they have problems. Susan Isaacs, A. S. Neill, and many of those who run schools for the maladjusted, have shown how a teacher's understanding of the deeper desires and fantasies of children can help them to fulfil their personal potential. The factor-analytical approach, on the other hand, has concentrated much more on the relationship between personality dimensions and the learner's academic attainment.

Asher Cashdan

4 An outline of psycho-analysis

S. Freud

Preface

The aim of this brief work is to bring together the tenets of psycho-analysis and to state them, as it were, dogmatically–in the most concise form and in the most unequivocal terms. Its intention is naturally not to compel belief or to arouse conviction.

The teachings of psycho-analysis are based on an incalculable number of observations and experiences, and only someone who has repeated those observations on himself and on others is in a position to arrive at a judgement of his own upon it.

I The psychical apparatus

Psycho-analysis makes a basic assumption, the discussion of which is reserved to philosophical thought but the justification for which lies in its results. We know two kinds of things about what we call our psyche (or mental life): firstly, its bodily organ and scene of action, the brain (or nervous system); and, on the other hand, our acts of consciousness, which are immediate data and cannot be further explained by any sort of description. Everything that lies between is unknown to us, and the data do not include any direct relation between these two terminal points of our knowledge. If it existed, it would at the most afford an exact localization of the processes of consciousness and would give us no help towards understanding them.

Our two hypotheses start out from these ends or beginnings of our knowledge. The first is concerned with localization.[1] We assume that mental life is the function of an apparatus to which we ascribe the characteristics of being extended in space and of being made up of several portions–which we imagine, that is, as resembling a telescope or microscope or something of the kind. Notwithstanding some earlier attempts in the same direction, the consistent working-out of a conception such as this is a scientific novelty.

---

[1] [The second is stated on p. 46 below.]

Source: FREUD, S. (1940) ed. STRACHEY, J. (1964) *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, Vol. 23: *An Outline of Psycho-analysis*. London: Hogarth Press and Institute of Psycho-analysis, 144-71.

We have arrived at our knowledge of this psychical apparatus by studying the individual development of human beings. To the oldest of these psychical provinces or agencies we give the name of *id*. It contains everything that is inherited, that is present at birth, that is laid down in the constitution–above all, therefore, the instincts, which originate from the somatic organization and which find a first psychical expression here [in the id] in forms unknown to us.[2]

Under the influence of the real external world around us, one portion of the id has undergone a special development. From what was originally a cortical layer, equipped with the organs for receiving stimuli and with arrangements for acting as a protective shield against stimuli, a special organization has arisen which henceforward acts as an intermediary between the id and the external world. To this region of our mind we have given the name of *ego*.

Here are the principal characteristics of the ego. In consequence of the pre-established connection between sense perception and muscular action, the ego has voluntary movement at its command. It has the task of self-preservation. As regards *external* events, it performs that task by becoming aware of stimuli, by storing up experiences about them (in the memory), by avoiding excessively strong stimuli (through flight), by dealing with moderate stimuli (through adaptation) and finally by learning to bring about expedient changes in the external world to its own advantage (through activity). As regards *internal* events, in relation to the id, it performs that task by gaining control over the demands of the instincts, by deciding whether they are to be allowed satisfaction, by postponing that satisfaction to times and circumstances favourable in the external world or by suppressing their excitations entirely. It is guided in its activity by consideration of the tensions produced by stimuli, whether these tensions are present in it or introduced into it. The raising of these tensions is in general felt as *unpleasure* and their lowering as *pleasure*. It is probable, however, that what is felt as pleasure or unpleasure is not the *absolute* height of this tension but something in the rhythm of the changes in them. The ego strives after pleasure and seeks to avoid unpleasure. An increase in unpleasure that is expected and foreseen is met by a *signal of anxiety*; the occasion of such an increase, whether it threatens from without or within, is known as a *danger*. From time to time the ego gives up its connection with the external world and withdraws into the state of sleep, in which it makes far-reaching changes in its organization. It is to be inferred from the state of sleep that this organization consists in a particular distribution of mental energy.

The long period of childhood, during which the growing human being lives in dependence on his parents, leaves behind it as a precipitate the formation in his ego of a special agency in which this parental influence is prolonged. It has received the name of *super-ego*. In so far as this super-ego is differentiated from the ego or is opposed to it, it constitutes a third power which the ego must take into account.

---

[2] This oldest portion of the psychical apparatus remains the most important throughout life; moreover, the investigations of psycho-analysis started with it.

An action by the ego is as it should be if it satisfies simultaneously the demands of the id, of the super-ego and of reality—that is to say, if it is able to reconcile their demands with one another. The details of the relation between the ego and the super-ego become completely intelligible when they are traced back to the child's attitude to its parents. This parental influence of course includes in its operation not only the personalities of the actual parents but also the family, racial and national traditions handed on through them, as well as the demands of the immediate social *milieu* which they represent. In the same way, the super-ego, in the course of an individual's development, receives contributions from later successors and substitutes of his parents, such as teachers and models in public life of admired social ideals. It will be observed that, for all their fundamental difference, the id and the super-ego have one thing in common: they both represent the influences of the past—the id the influence of heredity, the super-ego the influence, essentially, of what is taken over from other people—whereas the ego is principally determined by the individual's own experience, that is by accidental and contemporary events.

This general schematic picture of a psychical apparatus may be supposed to apply as well to the higher animals which resemble man mentally. A super-ego must be presumed to be present wherever, as is the case with man, there is a long period of dependence in childhood. A distinction between ego and id is an unavoidable assumption. Animal psychology has not yet taken in hand the interesting problem which is here presented.

II The theory of the instincts

The power of the id expresses the true purpose of the individual organism's life. This consists in the satisfaction of its innate needs. No such purpose as that of keeping itself alive or of protecting itself from dangers by means of anxiety can be attributed to the id. That is the task of the ego, whose business it also is to discover the most favourable and least perilous method of obtaining satisfaction, taking the external world into account. The super-ego may bring fresh needs to the fore, but its main function remains the limitation of satisfactions.

The forces which we assume to exist behind the tensions caused by the needs of the id are called *instincts*. They represent the somatic demands upon the mind. Though they are the ultimate cause of all activity, they are of a conservative nature; the state, whatever it may be, which an organism has reached gives rise to a tendency to re-establish that state so soon as it has been abandoned. It is thus possible to distinguish an indeterminate number of instincts, and in common practice this is in fact done. For us, however, the important question arises whether it may not be possible to trace all these numerous instincts back to a few basic ones. We have found that instincts can change their aim (by displacement) and also that they can replace one another—the energy of one instinct passing over to another. This latter process is still insufficiently understood. After long hesitancies and vacillations we have decided to assume the existence of only two basic instincts, *Eros* and *the destructive instinct.* (The contrast between the instincts of self-preservation and the preservation of the species, as well as the contrast between ego-love and object-love, fall within Eros.) The aim of the first of these basic instincts is to establish ever greater unities and to preserve them thus—in short, to bind together; the aim of the second is, on the contrary, to undo connections and so to destroy things. In the case of the destructive instinct we may suppose that its final aim is to lead what is living into an inorganic state. For this reason we also call it the *death instinct.* If we assume that living things came later than inanimate ones and arose from them, then the death instinct fits in with the formula we have proposed to the effect that instincts tend towards a return to an earlier state. In the case of Eros (or the love instinct) we cannot apply this formula. To do so would presuppose that living substance was once a unity which had later been torn apart and was now striving towards re-union.

In biological functions the two basic instincts operate against each other or combine with each other. Thus, the act of eating is a destruction of the object with the final aim of incorporating it, and the sexual act is an act of aggression with the purpose of the most intimate union. This concurrent and mutually opposing action of the two basic instincts gives rise to the whole variegation of the phenomena of life. The analogy of our two basic instincts extends from the sphere of living things to the pair of opposing forces—attraction and repulsion—which rule in the inorganic world.

Modifications in the proportions of the fusion between the instincts have the most tangible results. A surplus of sexual aggressiveness will turn a lover into a sex-murderer, while a sharp diminution in the aggressive factor will make him bashful or impotent.

There can be no question of restricting one or the other of the basic instincts to one of the provinces of the mind. They must necessarily be met with everywhere. We may picture an initial state as one in which the total available energy of Eros, which henceforward we shall speak of as 'libido', is present in the still undifferentiated ego-id and serves to neutralize the destructive tendencies which are simultaneously present. (We are without a term analogous to 'libido' for describing the energy of the destructive instinct.) At a later stage it becomes relatively easy for us to follow the vicissitudes of the libido, but this is more difficult with the destructive instinct.

So long as that instinct operates internally, as a death instinct, it remains silent; it only comes to our notice when it is diverted outwards as an instinct of destruction. It seems to be essential for the preservation of the individual that this diversion should occur; the muscular apparatus serves this purpose. When the super-ego is established, considerable amounts of the aggressive instinct are fixated in the interior of the ego and operate there self-destructively. This is one of the dangers to health by which human beings are faced on their path to cultural development. Holding back aggressiveness is in general unhealthy and leads to illness (to mortification[3]). A person in a fit of rage will often demonstrate how the transition from aggressiveness that has been prevented to self-destructiveness is brought about by diverting the aggressiveness against himself: he tears his hair or beats his face with his fists, though he would evidently have preferred to apply this treatment to someone else. Some portion of self-destructiveness remains within, whatever the circumstances; till at last it succeeds in killing the individual, not, perhaps, until his libido has been used up or fixated in a disadvantageous way. Thus it may in general be suspected that the *individual* dies of his internal conflicts but that the *species* dies of its unsuccessful struggle against the external world if the latter changes in a fashion which cannot be adequately dealt with by the adaptations which the species has acquired.

It is hard to say anything of the behaviour of the libido in the id and in the super-ego. All that we know about it relates to the ego, in which at first the whole available quota of libido is stored up. We call this state absolute, primary *narcissism.* It lasts till the ego begins to cathect the ideas of objects with libido, to transform narcissistic libido into object-libido. Throughout the whole of life the ego remains the great reservoir from which libidinal cathexes are sent out to objects and into which they are also once more withdrawn, just as an amoeba behaves with its pseudopodia. It is only when a person is completely in love that the main quota of libido is transferred on to the object and the object to some extent takes the place of the ego. A characteristic of the libido which is important in life is its *mobility*, the facility with which it passes from one object to another. This must be contrasted with the *fixation* of the libido to particular objects, which often persists throughout life.

There can be no questions but that the libido has somatic sources, that it streams to the ego from various organs and parts of the body. This is most clearly seen in the case of that portion of the libido which, from its instinctual aim, is described as sexual excitation. The most prominent of the parts of the body from which this libido arises are known by the name of '*erotogenic zones*', though in fact the whole body is an erotogenic zone of this kind. The greater part of what we know about Eros–that is to say, about its exponent, the libido–has been gained from a study of the sexual function, which, indeed, on the prevailing view, even if not according to our theory, coincides with Eros. We have been able to form a picture of the way in which the sexual urge, which is destined to exercise a decisive influence on our life, gradually develops out of successive contributions from a number of component instincts, which represent particular erotogenic zones.

---

[3] ['*Kränkung*' means literally 'making ill'.]

III The development of the sexual function

According to the prevailing view human sexual life consists essentially in an endeavour to bring one's own genitals into contact with those of someone of the opposite sex. With this are associated, as accessory phenomena and introductory acts, kissing this extraneous body, looking at it and touching it. This endeavour is supposed to make its appearance at puberty–that is, at the age of sexual maturity–and to serve the purposes of reproduction. Nevertheless, certain facts have always been known which do not fit into the narrow framework of this view. (1) It is a remarkable fact that there are people who are only attracted by individuals of their own sex and by their genitals. (2) It is equally remarkable that there are people whose desires behave exactly like sexual ones but who at the same time entirely disregard the sexual organs or their normal use; people of this kind are known as 'perverts'. (3) And lastly it is a striking thing that some children (who are on that account regarded as degenerate) take a very early interest in their genitals and show signs of excitation in them.

It may well be believed that psycho-analysis provoked astonishment and denials when, partly on the basis of these three neglected facts, it contradicted all the popular opinions on sexuality. Its principal findings are as follows:

(*a*) Sexual life does not begin only at puberty, but starts with plain manifestations soon after birth.

(*b*) It is necessary to distinguish sharply between the concepts of 'sexual' and 'genital'. The former is the wider concept and includes many activities that have nothing to do with the genitals.

(*c*) Sexual life includes the function of obtaining pleasure from zones of the body–a function which is subsequently brought into the service of reproduction. The two functions often fail to coincide completely.

The chief interest is naturally focused on the first of these assertions, the most unexpected of all. It has been found that in early childhood there are signs of bodily activity to which only an ancient prejudice could deny the name of sexual and which are linked to psychical phenomena that we come across later in adult erotic life–such as fixation to particular objects, jealousy, and so on. It is further found, however, that these phenomena which emerge in early childhood form part of an ordered course of development, that they pass through a regular process of increase, reaching a climax towards the end of the fifth year, after which there follows a lull.

During this lull progress is at a standstill and much is unlearnt and there is much recession. After the end of this period of latency, as it is called, sexual life advances once more with puberty; we might say that it has a second efflorescence. And here we come upon the fact that the onset of sexual life is *diphasic*, that it occurs in two waves–something that is unknown except in man and evidently has an important bearing on hominization. It is not a matter of indifference that the events of this early period, except for a few residues, fall a victim to *infantile amnesia*. Our views on the aetiology of the neuroses and our technique of analytic therapy are derived from these conceptions; and our tracing of the developmental processes in this early period has also provided evidence for yet other conclusions.

The first organ to emerge as an erotogenic zone and to make libidinal demands on the mind is, from the time of birth onwards, the mouth. To begin with, all psychical activity is concentrated on providing satisfaction for the needs of that zone. Primarily, of course, this satisfaction serves the purpose of self-preservation by means of nourishment; but physiology should not be confused with psychology. The baby's obstinate persistence in sucking gives evidence at an early stage of a need for satisfaction which, though it originates from and is instigated by the taking of nourishment, nevertheless strives to obtain pleasure independently of nourishment and for that reason may and should be termed *sexual*.

During this oral phase sadistic impulses already occur sporadically along with the appearance of the teeth. Their extent is far greater in the second phase, which we describe as the sadistic-anal one, because satisfaction is then sought in aggression and in the excretory function. Our justification for including aggressive urges under the libido is based on the view that sadism is an instinctual fusion of purely libidinal and purely destructive urges, a fusion which thenceforward persists uninterruptedly.[4]

The third phase is that known as the phallic one, which is, as it were, a forerunner of the final form taken by sexual life and already much resembles it. It is to be noted that it is not the genitals of both sexes that play a part at this stage, but only the male ones (the phallus). The female genitals long remain unknown: in children's attempts to understand the sexual processes they pay homage to the venerable cloacal theory–a theory which has a genetic justification.[5]

With the phallic phase and in the course of it the sexuality of early childhood reaches its height and approaches its dissolution. Thereafter boys and girls have different histories. Both have begun to put their intellectual activity at the service of sexual researches; both start off from the premiss of the universal presence of the penis. But now the paths of the sexes diverge. The boy enters the Oedipus phase; he begins to manipulate his penis and simultaneously has phantasies of carrying out some sort of activity with it in relation to his mother, till, owing to the combined effect of a threat of castration and the sight of the absence of a penis in females, he experiences the greatest trauma of his life and this introduces the period of latency with all its consequences. The girl, after vainly attempting to do the same as the boy, comes to recognize her lack of a penis or rather the inferiority of her clitoris, with permanent effects on the development of her character; as a result of this first disappointment in rivalry, she often begins by turning away altogether from sexual life.

---

[4] The question arises whether the satisfaction of purely destructive instinctual impulses can be felt as pleasure, whether pure destructiveness without any libidinal admixture occurs. Satisfaction of the death instinct remaining in the ego seems not to produce feelings of pleasure, though masochism represents a fusion which is entirely analogous to sadism.

[5] The occurrence of early vaginal excitations is often asserted. But it is most probable that what is in question are excitations in the clitoris–that is, in an organ analogous to the penis. This does not invalidate our right to describe the phase as phallic.

It would be a mistake to suppose that these three phases succeed one another in a clear-cut fashion. One may appear in addition to another; they may overlap one another, may be present alongside of one another. In the early phases the different component instincts set about their pursuit of pleasure independently of one another; in the phallic phase there are the beginnings of an organization which subordinates the other urges to the primacy of the genitals and signifies the start of a coordination of the general urge towards pleasure into the sexual function. The complete organization is only achieved at puberty, in a fourth, genital phase. A state of things is then established in which (1) some earlier libidinal cathexes are retained, (2) others are taken into the sexual function as preparatory, auxiliary acts, the satisfaction of which produces what is known as fore-pleasure, and (3) other urges are excluded from the organization, and are either suppressed altogether (repressed) or are employed in the ego in another way, forming character-traits or undergoing sublimation with a displacement of their aims.

This process is not always performed faultlessly. Inhibitions in its development manifest themselves as the many sorts of disturbance in sexual life. When this is so, we find fixations of the libido to conditions in earlier phases, whose urge, which is independent of the normal sexual aim, is described as *perversion.* One such developmental inhibition, for instance, is homosexuality when it is manifest. Analysis shows that in every case a homosexual object-tie was present and in most cases persisted in a *latent* condition. The situation is complicated by the fact that as a rule the processes necessary for bringing about a normal outcome are not completely present or absent, but *partially* present, so that the final result remains dependent on these *quantitative* relations. In these circumstances the genital organization is, it is true, attained, but it lacks those portions of the libido which have not advanced with the rest and have remained fixated to pregenital objects and aims. This weakening shows itself in a tendency, if there is an absence of genital satisfaction or if there are difficulties in the real external world, for the libido to hark back to its earlier pregenital cathexes (*regression*).

During the study of the sexual functions we have been able to gain a first, preliminary conviction, or rather a suspicion, of two discoveries which will later be found to be important over the whole of our field. Firstly, the normal and abnormal manifestations observed by us (that is, the phenomenology of the subject) need to be described from the point of view of their dynamics and economics (in our case, from the point of view of the quantitative distribution of the libido). And secondly, the aetiology of the disorders which we study is to be looked for in the individual's developmental history—that is to say, in his early life.

IV Psychical qualities

I have described the structure of the psychical apparatus and the energies or forces which are active in it, and I have traced in a prominent example the way in which those energies (in the main, the libido) organize themselves into a physiological function which serves the purpose of the preservation of the species. There was nothing in all this to demonstrate the quite peculiar characteristic of what is psychical, apart, of course, from the empirical fact that this apparatus and these energies are the basis of the functions which we describe as our mental life. I will now turn to something which is uniquely characteristic of what is psychical, and which, indeed, according to a very widely held opinion, coincides with it to the exclusion of all else.

The starting-point for this investigation is provided by a fact without parallel, which defies all explanation or description—the fact of consciousness. Nevertheless, if anyone speaks of consciousness we know immediately and from our most personal experience what is meant by it. Many people, both inside and outside [psychological] science, are satisfied with the assumption that consciousness alone is psychical; in that case nothing remains for psychology but to discriminate among psychical phenomena between perceptions, feelings, thought-processes and volitions. It is generally agreed, however, that these conscious processes do not form unbroken sequences which are complete in themselves; there would thus be no alternative left to assuming that there are physical or somatic processes which are concomitant with the psychical ones and which we should necessarily have to recognize as more complete than the psychical sequences, since some of them would have conscious processes parallel to them but others would not. If so, it of course becomes plausible to lay the stress in psychology on these somatic processes, to see in *them* the true essence of what is psychical and to look for some other assessment of the conscious processes. The majority of philosophers, however, as well as many other people, dispute this and declare that the idea of something psychical being unconscious is self-contradictory.

But that is precisely what psycho-analysis is obliged to assert, and this is its second fundamental hypothesis [p. 37]. It explains the supposedly somatic concomitant phenomena as being what is truly psychical, and thus in the first instance disregards the quality of consciousness. It is not alone in doing this. Some thinkers (such as Theodor Lipps, for instance) have asserted the same thing in the same words; and the general dissatisfaction with the usual view of what is psychical has resulted in an increasingly urgent demand for the inclusion in psychological thought of a concept of the unconscious, though this demand has taken such an indefinite and obscure form that it could have no influence on science.

Now it would look as though this dispute between psycho-analysis and philosophy is concerned only with a trifling matter of definition–the question whether the name 'psychical' should be applied to one or another sequence of phenomena. In fact, however, this step has become of the highest significance. Whereas the psychology of consciousness never went beyond the broken sequences which were obviously dependent on something else, the other view, which held that the psychical is unconscious in itself, enabled psychology to take its place as a natural science like any other. The processes with which it is concerned are in themselves just as unknowable as those dealt with by other sciences, by chemistry or physics, for example; but it is possible to establish the laws which they obey and to follow their mutual relations and interdependences unbroken over long stretches–in short, to arrive at what is described as an 'understanding' of the field of natural phenomena in question. This cannot be effected without framing fresh hypotheses and creating fresh concepts; but these are not to be despised as evidence of embarrassment on our part but deserve on the contrary to be appreciated as an enrichment of science. They can lay claim to the same value as approximations that belongs to the corresponding intellectual scaffolding found in other natural sciences, and we look forward to their being modified, corrected and more precisely determined as further experience is accumulated and sifted. So too it will be entirely in accordance with out expectations if the basic concepts and principles of the new science (instinct, nervous energy, etc.) remain for a considerable time no less indeterminate than those of the older sciences (force, mass, attraction, etc.).

Every science is based on observations and experiences arrived at through the medium of our psychical apparatus. But since *our* science has as its subject that apparatus itself, the analogy ends here. We make our observations through the medium of the same perceptual apparatus, precisely with the help of the breaks in the sequence of 'psychical' events: we fill in what is omitted by making plausible inferences and translating it into conscious material. In this way we construct, as it were, a sequence of conscious events complementary to the unconscious psychical processes. The relative certainty of our psychical science is based on the binding force of these inferences. Anyone who enters deeply into our work will find that our technique holds its ground against any criticism.

In the course of this work the distinctions which we describe as psychical qualities force themselves on our notice. There is no need to characterize what we call 'conscious': it is the same as the consciousness of philosophers and of everyday opinion. Everything else psychical is in our view 'the unconscious'. We are soon led to make an important division in this unconscious. Some processes become conscious easily; they may then cease to be conscious, but can become conscious once more without any trouble: as people say, they can be reproduced or remembered. This reminds us that consciousness is in general a highly fugitive state. What is conscious is conscious only for a moment. If our perceptions do not confirm this, the contradiction is only an apparent one; it is explained by the fact that the stimuli which lead to perception may persist for considerable periods, so that meanwhile the perception of them may be repeated. The whole position is made clear in connection with the conscious perception of our thought-processes: these too may persist for some time, but they may just as well pass in a flash. Everything unconscious that behaves in this way, that can thus easily exchange the unconscious state for the conscious one, is therefore preferably described as 'capable of becoming conscious' or as *preconscious*. Experience has taught us that there is hardly a psychical process, however complicated it may be, which cannot on occasion remain preconscious, even though as a rule it will, as we say, push its way forward into consciousness. There are other psychical processes and psychical material which have no such easy access to becoming conscious but must be inferred, recognized and translated into conscious form in the manner described. For such material we reserve the name of the unconscious proper.

Thus we have attributed three qualities to psychical processes: they are either conscious, preconscious or unconscious. The division between the three classes of material which possess these qualities is neither absolute nor permanent. What is preconscious becomes conscious, as we have seen, without any assistance from us; what is unconscious can, through our efforts, be made conscious, and in the process we may have a feeling that we are often overcoming very strong resistances. When we attempt to do this with someone else, we should not forget that the conscious filling-in of the gaps in his perceptions–the construction we are presenting him with–does not mean as yet that we have made the unconscious material in question conscious to him. All that is true so far is that the material is present in him in two records,[6] once in the conscious reconstruction he has been given, and besides this in its original unconscious state. Our continued efforts usually succeed eventually in making this unconscious material conscious to him himself, as a result of which the two records are brought to coincide. The amount of effort we have to use, by which we estimate the resistance against the material becoming conscious, varies in magnitude in individual cases. For instance, what comes about in an analytic treatment as a result of our efforts can also occur spontaneously: material which is ordinarily unconscious can transform itself into preconscious material and then becomes conscious–a thing that happens to a large extent in psychotic states. From this we infer that the maintenance of certain internal resistances is a *sine qua non* of normality. A relaxation of resistances such as this, with a consequent pushing forward of unconscious material, takes place regularly in the state of sleep, and thus brings about a necessary precondition for the construction of dreams. Conversely, preconscious material can become temporarily inaccessible and cut off by resistances, as happens when something is temporarily forgotten or escapes the memory; or a preconscious thought can even be temporarily put back into the unconscious state, as seems to be a precondition in the case of jokes. We shall see that a similar transformation back of preconscious material or processes into the unconscious state plays a great part in the causation of neurotic disorders.

---

[6] [The German word translated by 'record' here is '*Fixierung*' used in exactly this sense in Chapter VII (B) of *The Interpretation of Dreams* (Freud, 1900. Vol. 5, 539).]

The theory of the three qualities of what is psychical, as described in this generalized and simplified manner, seems likely to be a source of limitless confusion rather than a help towards clarification. But it should not be forgotten that in fact it is not a theory at all but a first stock-taking of the facts of our observations, that it keeps as close to those facts as possible and does not attempt to explain them. The complications which it reveals may bring into relief the peculiar difficulties with which our investigations have to contend. It may be suspected, however, that we shall come to a closer understanding of this theory itself if we trace out the relations between the psychical qualities and the provinces or agencies of the psychical apparatus which we have postulated–though these relations too are far from being simple.

The process of something becoming conscious is above all linked with the perceptions which our sense organs receive from the external world. From the topographical point of view, therefore, it is a phenomenon which takes place in the outermost cortex of the ego. It is true that we also receive conscious information from the inside of the body–the feelings, which actually exercise a more peremptory influence on our mental life than external perceptions; moreover, in certain circumstances the sense organs themselves transmit feelings, sensations of pain, in addition to the perceptions specific to them. Since, however, these sensations (as we call them in contrast to conscious perceptions) also emanate from the terminal organs and since we regard all these as prolongations or offshoots of the cortical layer, we are still able to maintain [this] assertion. The only distinction would be that, as regards the terminal organs of sensation and feeling, the body itself would take the place of the external world.

Conscious processes on the periphery of the ego and everything else in the ego unconscious–such would be the simplest state of affairs that we might picture. And such may in fact be the state that prevails in animals.

But in men there is an added complication through which internal processes in the ego may also acquire the quality of consciousness. This is the work of the function of speech, which brings material in the ego into a firm connection with mnemic residues of visual, but more particularly of auditory, perceptions. Thenceforward the perceptual periphery of the cortical layer can be excited to a much greater extent from inside as well, internal events such as passages of ideas and thought-processes can become conscious, and a special device is called for in order to distinguish between the two possibilities–a device known as *reality-testing.* The equation 'perception = reality (external world)' no longer holds. Errors, which can now easily arise and do so regularly in dreams, are called *hallucinations.*

The inside of the ego, which comprises above all the thought-processes, has the quality of being preconscious. This is characteristic of the ego and belongs to it alone. It would not be correct, however, to think that connection with the mnemic residues of speech is a necessary precondition of the preconscious state. On the contrary, that state is independent of a connection with them, though the presence of that connection makes it safe to infer the preconscious nature of a process. The preconscious state, characterized on the one hand by having access to consciousness and on the other hand by its connection with the speech-residues, is nevertheless something peculiar, the nature of which is not exhausted by these two characteristics. The evidence for this is the fact that large portions of the ego, and particularly of the super-ego, which cannot be denied the characteristic of preconsciousness, none the less remain for the most part unconscious in the phenomenological sense of the word. We do not know why this must be so. We shall attempt presently to attack the problem of the true nature of the preconscious.

The sole prevailing quality in the id is that of being unconscious. Id and unconscious are as intimately linked as ego and preconscious: indeed, in the former case the connection is even more exclusive. If we look back at the developmental history of an individual and of his psychical apparatus, we shall be able to perceive an important distinction in the id. Originally, to be sure, everything was id; the ego was developed out of the id by the continual influence of the external world. In the course of this slow development certain of the contents of the id were transformed into the preconscious state and so taken into the ego; others of its contents remained in the id unchanged, as its scarcely accessible nucleus. During this development, however, the young and feeble ego put back into the unconscious state some of the material it had already taken in, dropped it, and behaved in the same way to some fresh impressions which it *might* have taken in, so that these, having been rejected, could leave a trace only in the id. In consideration of its origin we speak of this latter portion of the id as *the repressed.* It is of little importance that we are not always able to draw a sharp line between these two categories of contents in the id. They coincide approximately with the distinction between what was innately present originally and what was acquired in the course of the ego's development.

Having now decided upon the topographical dissection of the psychical apparatus into an ego and an id, with which the difference in quality between preconscious and unconscious runs parallel, and having agreed that this quality is to be regarded only as an *indication* of the difference and not as its essence, a further question faces us. What, if this is so, is the true nature of the state which is revealed in the id by the quality of being unconscious and in the ego by that of being preconscious and in what does the difference between them consist?

But of that we know nothing. And the profound obscurity of the background of our ignorance is scarcely illuminated by a few glimmers of insight. Here we have approached the still shrouded secret of the nature of the psychical. We assume, as other natural sciences have led us to expect, that in mental life some kind of energy is at work; but we have nothing to go upon which will enable us to come nearer to a knowledge of it by analogies with other forms of energy. We seem to recognize that nervous or psychical energy occurs in two forms, one freely mobile and another, by comparison, bound; we speak of cathexes and hypercathexes of psychical material, and even venture to suppose that a hypercathexis brings about a kind of synthesis of different processes–a synthesis in the course of which free energy is transformed into bound energy. Further than this we have not advanced. At any rate, we hold firmly to the view that the distinction between the unconscious and the preconscious state lies in dynamic relations of this kind, which would explain how it is that, whether spontaneously or with our assistance, the one can be changed into the other.

Behind all these uncertainties, however, there lies one new fact, whose discovery we owe to psycho-analytic research. We have found that processes in the unconscious or in the id obey different laws from those in the preconscious ego. We name these laws in their totality the *primary process*, in contrast to the *secondary process* which governs the course of events in the preconscious, in the ego. In the end, therefore, the study of psychical qualities has after all proved not unfruitful.

V Dream-interpretation as an illustration

An investigation of normal, stable states, in which the frontiers of the ego are safeguarded against the id by resistances (anticathexes) and have held firm, and in which the super-ego is not distinguished from the ego, because they work together harmoniously–an investigation of that kind would teach us little. The only thing that can help us are states of conflict and uproar, when the contents of the unconscious id have a prospect of forcing their way into the ego and into consciousness and the ego puts itself once more on the defensive against this invasion. It is only under these conditions that we can make such observations as will confirm or correct our statements about the two partners. Now, our nightly sleep is precisely a state of this sort, and for that reason psychical activity during sleep, which we perceive as dreams, is our most favourable object of study. In that way, too, we avoid the familiar reproach that we base our constructions of normal mental life on pathological findings; for dreams are regular events in the life of a normal person, however much their characteristics may differ from the productions of our waking life. Dreams, as everyone knows, may be confused, unintelligible or positively nonsensical, what they say may contradict all that we know of reality, and we behave in them like insane people, since, so long as we are dreaming, we attribute objective reality to the contents of the dream.

We find our way to the understanding ('interpretation') of a dream by assuming that what we recollect as the dream after we have woken up is not the true dream-process but only a *façade* behind which that process lies concealed. Here we have our distinction between the *manifest* content of a dream and the *latent* dream-thoughts. The process which produces the former out of the latter is described as the *dream-work*. The study of the dream-work teaches us by an excellent example the way in which unconscious material from the id (originally unconscious and repressed unconscious alike) forces its way into the ego, becomes preconscious and, as a result of the ego's opposition, undergoes the changes which we know as *dream-distortion*. There are no features of a dream which cannot be explained in this way.

It is best to begin by pointing out that the formation of a dream can be provoked in two different ways. Either, on the one hand, an instinctual impulse which is ordinarily suppressed (an unconscious wish) finds enough strength during sleep to make itself felt by the ego, or, on the other hand, an urge left over from waking life, a preconscious train of thought with all the conflicting impulses attached to it, finds reinforcement during sleep from an unconscious element. In short, dreams may arise either from the id or from the ego. The mechanism of dream-formation is in both cases the same and so also is the necessary dynamic precondition. The ego gives evidence of its original derivation from the id by occasionally ceasing its functions and allowing a reversion to an earlier state of things. This is logically brought about by its breaking off its relations with the external world and withdrawing its cathexes from the sense organs. We are justified in saying that there arises at birth an instinct to return to the intra-uterine life that has been abandoned—an instinct to sleep. Sleep is a return of this kind to the womb. Since the waking ego governs motility, that function is paralysed in sleep, and accordingly a good part of the inhibitions imposed on the unconscious id become superfluous. The withdrawal or reduction of these 'anticathexes' thus allows the id what is now a harmless amount of liberty.

The evidence of the share taken by the unconscious id in the formation of dreams is abundant and convincing. (*a*) Memory is far more comprehensive in dreams than in waking life. Dreams bring up recollections which the dreamer has forgotten, which are inaccessible to him when he is awake. (*b*) Dreams make an unrestricted use of linguistic symbols, the meaning of which is for the most part unknown to the dreamer. Our experience, however, enables us to confirm their sense. They probably originate from earlier phases in the development of speech. (*c*) Memory very often reproduces in dreams impressions from the dreamer's early childhood of which we can definitely assert not only that they had been forgotten but that they had become unconscious owing to repression. That explains the help–usually indispensable–given us by dreams in the attempts we make during the analytic treatment of neuroses to reconstruct the dreamer's early life. (*d*) Furthermore, dreams bring to light material which cannot have originated either from the dreamer's adult life or from his forgotten childhood. We are obliged to regard it as part of the *archaic heritage* which a child brings with him into the world, before any experience of his own, influenced by the experiences of his ancestors. We find the counterpart of this phylogenetic material in the earliest human legends and in surviving customs. Thus dreams constitute a source of human prehistory which is not to be despised.

But what makes dreams so invaluable in giving us insight is the circumstance that, when the unconscious material makes its way into the ego, it brings its own modes of working along with it. This means that the preconscious thoughts in which the unconscious material has found its expression are handled in the course of the dream-work as though they were unconscious portions of the id; and, in the case of the alternative method of dream-formation, the preconscious thoughts which have obtained reinforcement from an unconscious instinctual impulse are brought down to the unconscious state. It is only in this way that we learn the laws which govern the passage of events in the unconscious and the respects in which they differ from the rules that are familiar to us in waking thought. Thus the dream-work is essentially an instance of the unconscious working-over of preconscious thought-processes. To take an analogy from history: invading conquerors govern a conquered country, not according to the judicial system which they find in force there, but according to their own. It is, however, an unmistakable fact that the outcome of the dream-work is a compromise. The ego-organization is not yet paralysed, and its influence is to be seen in the distortion imposed on the unconscious material and in what are often very ineffective attempts at giving the total result a form not too unacceptable to the ego (*secondary revision*). In our analogy this would be an expression of the continued resistance of the defeated people.

The laws that govern the passage of events in the unconscious, which come to light in this manner, are remarkable enough and suffice to explain most of what seems strange to us about dreams. Above all there is a striking tendency to *condensation*, an inclination to form fresh unities out of elements which in our waking thought we should certainly have kept separate. As a consequence of this, a single element of the manifest dream often stands for a whole number of latent dream-thoughts as though it were a combined allusion to all of them; and in general the compass of the manifest dream is extraordinarily small in comparison with the wealth of material from which it has sprung. Another peculiarity of the dream-work, not entirely independent of the former one, is the ease with which psychical intensities[7] (cathexes) are *displaced* from one element to another, so that it often happens that an element which was of little importance in the dream-thoughts appears as the clearest and accordingly most important feature of the manifest dream, and, vice versa, that essential elements of the dream-thoughts are represented in the manifest dream only by slight allusions. Moreover, as a rule the existence of quite insignificant points in common between two elements is enough to allow the dream-work to replace one by the other in all further operations. It will easily be imagined how greatly these mechanisms of condensation and displacement can increase the difficulty of interpreting a dream and of revealing the relations between the manifest dream and the latent dream-thoughts. From the evidence of the existence of these two tendencies to condensation and displacement our theory infers that in the unconscious id the energy is in a freely mobile state and that the id sets more store by the possibility of discharging quantities of excitation than by any other consideration;[8] and our theory makes use of these two peculiarities in defining the character of the primary process we have attributed to the id.

The study of the dream-work has taught us many other characteristics of the processes in the unconscious which are as remarkable as they are important; but we must only mention a few of them here. The governing rules of logic carry no weight in the unconscious; it might be called the Realm of the Illogical. Urges with contrary aims exist side by side in the unconscious without any need arising for an adjustment between them. Either they have no influence whatever on each other, or, if they have, no decision is reached, but a compromise comes about which is nonsensical since it embraces mutually incompatible details. With this is connected the fact that contraries are not kept apart but treated as though they were identical, so that in the manifest dream any element may also have the meaning of its opposite. Certain philologists have found that the same held good in the most ancient languages and that contraries such as 'strong-weak', 'light-dark' and 'high-deep' were originally expressed by the same roots, until two different modifications of the primitive word distinguished between the two meanings. Residues of this original double meaning seem to have survived even in a highly developed language like Latin in its use of words such as '*altus*' ('high' and deep') and '*sacer*' ('sacred' and 'infamous').

---

[7] [A term very often used by Freud from the earliest times as an equivalent to psychical energy.]

[8] An analogy may be seen in the behaviour of a non-commissioned officer who accepts a reprimand from his superior in silence but vents his anger on the first innocent private he comes across.

In view of the complication and ambiguity of the relations between the manifest dream and the latent content lying behind it, it is of course justifiable to ask how it is at all possible to deduce the one from the other and whether all we have to go on is a lucky guess, assisted perhaps by a translation of the symbols that occur in the manifest dream. It may be said in reply that in the great majority of cases the problem can be satisfactorily solved, but only with the help of the associations provided by the dreamer himself to the elements of the manifest content. Any other procedure is arbitrary and can yield no certain result. But the dreamer's associations bring to light intermediate links which we can insert in the gap between the two [between the manifest and latent content] and by aid of which we can reinstate the latent content of the dream and 'interpret' it. It is not to be wondered at if this work of interpretation (acting in a direction opposite to the dream-work) fails occasionally to arrive at complete certainty.

It remains for us to give a dynamic explanation of why the sleeping ego takes on the task of the dream-work at all. The explanation is fortunately easy to find. With the help of the unconscious, every dream that is in process of formation makes a demand upon the ego—for the satisfaction of an instinct, if the dream originates from the id; for the solution of a conflict, the removal of a doubt or the forming of an intention, if the dream originates from a residue of preconscious activity in waking life. The sleeping ego, however, is focused on the wish to maintain sleep; it feels this demand as a disturbance and seeks to get rid of the disturbance. The ego succeeds in doing this by what appears to be an act of compliance: it meets the demand with what is in the circumstances a harmless *fulfilment of a wish* and so gets rid of it. This replacement of the demand by the fulfilment of a wish remains the essential function of the dream-work. It may perhaps be worth while to illustrate this by three simple examples—a hunger dream, a dream of convenience and a dream prompted by sexual desire. A need for food makes itself felt in a dreamer during his sleep: he has a dream of a delicious meal and sleeps on. The choice, of course, was open to him either of waking up and eating something or of continuing his sleep. He decided in favour of the latter and satisfied his hunger by means of the dream—for the time being, at all events, for if his hunger had persisted he would have had to wake up nevertheless. Here is the second example. A sleeper had to wake up so as to be in time for his work at the hospital. But he slept on, and had a dream that he was already at the hospital—but as a patient, who has no need to get up. Or again, a desire becomes active during the night for the enjoyment of a forbidden sexual object, the wife of a friend of the sleeper. He has a dream of sexual intercourse—not, indeed, with this person but with someone else of the same name to whom he is in fact indifferent; or his struggle against the desire may find expression in his mistress remaining altogether anonymous.

Naturally, every case is not so simple. Especially in dreams which have originated from undealt-with residues of the previous day, and which have only obtained an unconscious reinforcement during the state of sleep, it is often no easy task to uncover the unconscious motive force and its wish-fulfilment; but we may assume that it is always there. The thesis that dreams are the fulfilments of wishes will easily arouse scepticism when it is remembered how many dreams have an actually distressing content or even wake the dreamer in anxiety, quite apart from the numerous dreams without any definite feeling-tone. But the objection based on anxiety dreams cannot be sustained against analysis. It must not be forgotten that dreams are invariably the product of a conflict, that they are a kind of compromise-structure. Something that is a satisfaction for the unconscious id may for that very reason be a cause of anxiety for the ego.

As the dream-work proceeds, sometimes the unconscious will press forward more successfully and sometimes the ego will defend itself with greater energy. Anxiety dreams are mostly those whose content has undergone the least distortion. If the demand made by the unconscious is too great for the sleeping ego to be in a position to fend it off by the means at its disposal, it abandons the wish to sleep and returns to waking life. We shall be taking every experience into account if we say that a dream is invariably an *attempt* to get rid of a disturbance of sleep by means of a wish-fulfilment, so that the dream is a guardian of sleep. The attempt may succeed more or less completely; it may also fail, and in that case the sleeper wakes up, apparently woken precisely by the dream. So, too, there are occasions when that excellent fellow the night-watchman, whose business it is to guard the little township's sleep, has no alternative but to sound the alarm and waken the sleeping townspeople.

I will close this discussion with a comment which will justify the length of time I have spent on the problem of the interpretation of dreams. Experience has shown that the unconscious mechanisms which we have come to know from our study of the dream-work and which gave us the explanation of the formation of dreams also help us to understand the puzzling symptoms which attract our interest to neuroses and psychoses. A conformity of such a kind cannot fail to excite high hopes in us.

Reference

FREUD, S. (1900) ed. STRACHEY, J. (1953) *The Standard Edition of the Complete Psychological Works of Sigmund Freud,* Vols. 4 and 5: *The Interpretation of Dreams.* London: Hogarth Press and Institute of Psycho-analysis.

5 Empirical studies of the Oedipus and castration complexes

P. Kline

[This extract forms part of a chapter in Paul Kline's book, *Fact and Fantasy in Freudian Theory,* in which he reviews a series of studies on the Oedipus and castration complexes.]

The work of Kuppuswamy (1949)

Kuppuswamy attempted to investigate the incidence of the Oedipus complex by sending questionnaires asking which child parents preferred and which parent children preferred. This study merits only the briefest consideration since it is highly unlikely that a questionnaire could tap the Oedipus complex which was only discovered or inferred from lengthy free associations. In summary, Kuppuswamy found that fathers and mothers prefer sons and that boys and girls prefer mothers. This result was regarded as partial confirmation of the Oedipus complex.

Conclusions and comments

First of all, simple questions of preference are not relevant to the Oedipus complex. They ignore defence mechanisms such as reaction-formation and sublimation–also part of Freudian theory and not to be ignored since, according to the theory, the Oedipus complex in boys is 'broken upon the rocks' of the castration complex–that is, it is repressed and unconscious. In addition, problems of social desirability of responses are inherent in the questionnaire method. This study, therefore, may be dismissed as of little value. It is a good example of one of the reasons psychoanalysts ignore psychological research into psychoanalytic theory: the research concerns itself with psychoanalytic concepts so grossly distorted as to be unrecognizable.

The work of McElroy (1950)

McElroy discussed three methods of testing the Oedipus complex. Unfortunately he gives no report of the results thus obtained, but the paper is worth brief examination because it provides a useful background to the

---

Source: KLINE, P. (1972) *Fact and Fantasy in Freudian Theory.* London: Methuen Educational, 97-110.

methodology of empirical investigations. It indicates immediately, for example, the shortcomings of the work of Kuppuswamy (1949). A short description of the methods is set out below.

1. A questionnaire is constructed to deal with all possible aspects of the physical appearance of women and this includes the use of drawings. Half a sample of young married men complete this questionnaire first concerning their mothers and then their wives. The other half reverse this order. The intensity of the Oedipus complex (or its presence) is measured by the correlation between the two profiles.
2. Photographs of mothers and wives are used. Three copies of the photograph of each wife are necessary. Then cards are constructed with four photographs. The mother of subject X is in the left hand corner. One of the other three pictures is the wife of subject X, the other two photographs are the wives of two other subjects selected at random. Subjects in the experiment are told the aim of this is to assess the likeness in appearance of mothers and daughters. Since selection of wife and mother will occur by chance in the proportion of one in three, an index of Oedipal strength may easily be estimated.
3. Parent preference in children. Children use pastel colours on grey paper. Then they draw their parents in pastel. Finally they are asked to name their favourite colour. The colour of each parent is then noted.

Comments

These methods may be easily elaborated. For example, a more accurate way to find the child's favourite colour would be by the psychophysical method of paired comparisons (see Woodworth and Schlosberg, 1955) where every colour is compared with each other. Again, in method two, care would have to be taken that the photographs were of a similar quality. Nevertheless it is clear that these three methods are all viable ways of attempting to assess both the strength and the existence of the Oedipus complex as it is conceived in Freudian theory.

The work of Friedman

Friedman (1950b) summarized some results of his study of the Oedipus complex in the *American Psychologist.* This paper was, in fact, part of his doctoral dissertation (Friedman, 1950a) and it was more comprehensively written up in a later paper (Friedman, 1952). These three reports will therefore not receive separate treatment.

For his sample he used 305 normal children, 5-16 years old. Age groups were formed to two years span. In each group there were 26 male and 26 female children.

Hypothesis
A. Castration complex

At the age of five years fears of castration will be great. During latency these fears will subside but by puberty they will be great again.

Method
Three unfinished 'castration fables' were presented to the children to complete. It was hypothesized that there would be *fewer* mentions of loss of tails, etc., at five years and thirteen years than at other ages. Fewer castration stories were expected when castration anxiety was high because it had previously been shown that reaction-time to stories with no cutting in them was longer.

The fables used by Friedman (1952) were those originally used by Duss (1940) and translated by Despert (1946). The castration-complex fable is known as 'The Elephant'. In this a child comes into a playroom and finds his favourite toy, an elephant, broken. The subject is asked to say what he thinks is wrong with it.

Results
Boys. Significantly more had stories without cutting at ages of five and thirteen years than at other ages. Freudian theory was, therefore, supported. *Girls.* There was no difference between Oedipal ages and latency. Overall 75 per cent of children had castration fears. This is interesting in the light of Sears' (1943) claim that Freud grossly overestimated the incidence of castration fears.

B. The Oedipus complex

Friedman reviewed the earlier studies of the Oedipus complex, many of which have been discussed by Sears (1943) who concluded that Freud's notion of the Oedipus complex was a grotesquerie. Most of these studies simply asked the children which parent they preferred. As has already been pointed out, these questions are irrelevant to the Freudian concept of the Oedipus complex which is repressed and therefore unconscious. The fact that the majority of these investigations failed to support the theory is, therefore, irrelevant.

Method
1. There were two fables for the children to complete: a child enjoys himself with one parent then meets the other. Why does the face of 'at home' parent look different? Then the same fable, with parental roles reversed, has to be completed by the children.
2. Three TAT[1]-like pictures were shown:

---

[1] [Thematic Apperception Test.]

(*a*) *Boys*: father and son are standing by the stairs and near them is a toy waggon. *Girls*: father and daughter by stairs, with a toy doll.
(*b*) Mother and child.
(*c*) Father and mother.

Data

1. Direct questions as to parental preference, as in other investigations.
2. Analysis of themes to pictures.
3. Analysis of movement tendencies and reactions to common psychoanalytic symbols.

Results

1. *Direct questions.* There were no significant differences between the sexes. Each sex liked each parent equally. This, therefore, follows the results of other similar investigations and fails to 'support' Freudian theory.
2. *Incomplete fables.* There were more negative endings to the fable where a child met the parent of his own sex than to the other fable. This fits Freudian theory in that the child would be expected to project his own feelings on to the parent. The difference was significant at the 0.05 level. Method 2 therefore supports Freudian theory.
3. *Themes*
(*a*) Boys had more conflict themes to father figures as did girls to mother figures. The difference was significant at the 0.001 level. Freudian theory is again supported.
(*b*) More girls fantasized that father-figure takes some positive action to toy than boys. The difference was significant at the 0.001 level and confirms Freudian theory.
(*c*) More girls than boys (sig. 0.001 level) fantasized that the male figure mounts the stairs and enters the room. Again this strongly supports Freudian theory in that mounting stairs in psychoanalytic theory symbolizes sexual intercourse and entering a room symbolizes inserting the penis into the vagina.

As Friedman (1952) points out, all methods other than the one most commonly used, which has been shown to be faulty, support the Freudian theory of the Oedipus complex.

Comments and conclusions

These results are so clear-cut and the hypotheses tested are so clearly derivable from the psychoanalytic theory that little comment is necessary. The large sample, 307, means that the investigation cannot be attacked on grounds of sample size. The fact that the pictures and fables constitute a projective technique does not mean that the results must be regarded as untrustworthy. The projective techniques used here, as is the case with the

Blacky Pictures, have a clear rationale rooted in psychoanalytic theory. The results hypothesized depend upon this theory, not on the theory of the test itself as is the case with the Rorschach and the Szondi. The last finding concerning the mounting of the staircase is the most extraordinary. Only on Freudian Oedipal theory could such a finding have been predicted. This result supports, also, Freudian claims about symbolization.

It must be concluded therefore that this study by Friedman (1952) has made three main points:

1. It has empirically demonstrated the inadequacy of the questionnaire approach to investigations of the Oedipus and castration complexes.
2. It has provided good evidence for the existence of a castration complex in boys.
3. It has provided good evidence for the existence of the Oedipus complex in boys and girls.

The work of Michal-Smith, Hammer and Spitz (1951)

This was a case study of a Negro boy whose Oedipal desires were very close to consciousness. The Oedipal dimension of the Blacky Pictures (see Appendix, p. 70) agreed with the psychiatric diagnosis and showed him to be highly disturbed about such conflicts. This is, at least, an objective verification of clinical data and as such supports the existence of the Oedipus complex.

Survey by Roheim (1952)

Roheim reviewed some anthropological evidence concerning the Oedipus complex, in an attempt to show that the Oedipus complex occurs even in those societies where free sexual play is allowed for children. Although the evidence quoted is not objective in that no standardized measurements were used, the data are pertinent to the Oedipus complex, being myths, dreams and characteristic doll-play. Thus, according to Berndt and Berndt (1951) there is a myth among the Alaiva in Arnhem Land about Kadjari, who killed his son because the boy wanted to lie with his mother. Both now live on in a heavenly constellation. Yet among this tribe children are given complete sexual freedom. Among the Lesu in New Ireland (Powdermaker, 1933) Oedipal dreams, of which the men are much ashamed, are common. Henry and Henry (1944), in their study of the Pilaga Indians, claimed that the doll-play of the children showed clear indications of the Oedipus complex. Finally, the Fan of West Africa (Trilles, 1912) have a myth of Crocodile, son of Crocodile, who kills his crocodile father. Yet in these three societies there is freedom of sexual play for the children.

Now it is not argued that these anthropological findings are objective evidence for the existence of an Oedipus complex. However, it would be wrong to ignore them totally, because the power of a theory depends upon its ability to comprehend a wide variety of evidence and it is in this that psychoanalytic theory is outstanding. These anthropological findings are examples of such a comprehensive power.

The work of Schwartz

Schwartz (1955) developed from the TAT and validated a measure of castration anxiety. Using this measure, Schwartz (1956) then carried out an empirical study of two Freudian hypotheses concerning castration anxiety.

The measure of castration anxiety (Schwartz, 1955)

Method and sample
Three groups of students, eighteen in each group. There were three treatments–different films were shown. The *castration group* were shown a film of sub-incision rites. The *anxiety group* were shown a film about Negro deprivation. The *control group* saw a Chaplin comedy film.

The TAT variables. The TAT cards were subjected to a content analysis. Interrater correlation was high, 0.80. Ten variables were extracted and examined. Validation was demonstrated if the castration group was discriminated from the other two groups. A list of the variables is set out below: whether they were valid or not is also indicated.

List of variables and results	*Valid*
1. Genital injury or loss	No
2. Loss or damage to other parts of the body	No
3. Damage to body image or symbols (e.g. guns)	Yes
4. Sexual inadequacy	No
5. Personal inadequacy	Yes
6. Repetitive attempts at mastery (e.g. risk-taking)	Yes
7. Intrapsychic threat: guilt, expectation of punishment	Yes
8. Extrapsychic threat: retaliation from external sources	Yes
9. Loss of cathected objects	Yes
10. Total score 1-9	Yes

As was pointed out by Schwartz (1955), Variables 1, 2, and 4 probably failed because they were simply too obvious for normal subjects.

Comments and conclusions

The fact that these variables, derived from the TAT, can discriminate the castration group from the other two groups is good evidence for their validity as measures of castration anxiety. Since one of the groups was an anxiety group, this discriminative ability cannot be due to their being made anxious by pictures of sub-incision. This validity is, on its own, support for the Freudian concept of castration anxiety. If no such anxiety existed the first two groups would not score differently.

Schwartz (1956)

With this TAT measure of castration anxiety, Schwartz (1956) made further investigations of the castration complex:

Sample. Twenty homosexual male students compared with a control group of twenty normal students, matched for age, education and socio-economic status.
Tests. Variables as in the previous study.
Results. Variables: 4–sexual inadequacy, 7–intrapsychic threat, 8–extrapsychic threat, and 11–a new variable based on the formal characteristics of the story (mis-spelling and discontinuity, for example) discriminated the groups in the predicted direction at the 1 per cent level by chi-square–as did the total score. Variable 5, personal inadequacy, discriminated at the 5 per cent level by chi-square. Similar but less significant results were obtained using the Link-Wallace statistic (Lindzey, 1954). Variables 6, 2, 3 and 9 failed to discriminate. Schwartz (1956) argues that these results support the psychoanalytic claim that castration anxiety can create homosexuality.

Comments and conclusions

Four of the variables fail to support the Freudian theory. Furthermore, variables 4 and 11, which discriminate the homosexuals, were not, in the previous study (Schwartz, 1955), related to castration anxiety. Thus only three of the original scores discriminate the groups. This cannot be regarded as good support for Freudian theory. Clearly a replication of this study is needed with more substantial samples. Nevertheless it is slight evidence for the psychoanalytic hypothesis that homosexuality is linked with castration anxiety. It is to be noted that such evidence as this inherently cannot demonstrate a *causal* link.

Schwartz (1956) also investigated sex differences in castration anxiety:

Hypothesis. Men have greater castration anxiety than women.
Sample. Twenty matched pairs of males and females.
Results. Variables 5 and 6–personal inadequacy and repetitive attempts at mastery–discriminated the groups at the 5 per cent level, as did the total score. The other differences were non-significant.

Comments and conclusions

The Freudian hypothesis receives some support although it is clear that there is a considerable over-lap of distributions. It is interesting that variable 5, which failed to discriminate before, was this time successful. From this study the Freudian theory that men have greater castration anxiety than women is supported to a limited extent.

Conclusions from the studies by Schwartz (1955, 1956)

These studies provide good evidence for the existence of castration anxiety and a useful measure of it. However, the investigations with this TAT device give only limited support to the notion that homosexuality is linked with castration anxiety and that men have greater castration anxiety than women. Clearly more studies are needed with these variables and a factor analysis of them would enable weighted scores from each to be used, perhaps allowing for greater discriminative power. In brief the results may be thus summarized:

1. The existence of castration anxiety: good support.
2. The link between homosexuality and castration anxiety: limited support.
3. Greater castration anxiety among men than women: limited support.

The work of Biddle (1957)

Biddle regressed 100 subjects back to three years old under hypnosis and then investigated their Oedipal fantasies. The major finding was that all children strive for union with both parents. The conclusion, therefore, was that the Oedipus complex was not supported.

Comments and conclusions

This is an interesting study in that hypnosis was the original technique by which unconscious mental activity was uncovered in the beginnings of psychoanalysis. Only later was free association preferred, partly because Freud found difficulty in hypnotizing patients. However, the assumption that hypnotic regression to any age gives a true representation of what in fact occurred at that age is by no means proven. Furthermore, the choice of three years of age to investigate the Oedipus complex is curious. Yet another objection to the study involves the lack of quantification of the data and their interpretative nature. Thus the assertion that the concept of the Oedipus complex was not confirmed by this study is not considered of great importance in the empirical study of the problem. This investigation is best ignored.

The work of Rabin (1958)

Rabin investigated some psychosexual differences between Kibbutz and non-Kibbutz Israeli boys:

Sample. Twenty-seven boys reared in a Kibbutz. Mean age, 10 years 2 months. Twenty-seven boys reared in their families. Mean age, 10 years 2 months.
Test. The Blacky Pictures (Blum, 1949).
Hypotheses
1. Fewer Kibbutz-reared children will exhibit Oedipal intensity compared with the family children.
2. Less identification with the father will be noted among the Kibbutz children compared with the family children.
Results. These hypotheses were supported. Among the Kibbutz boys there was less Oedipal intensity and less identification with the father.

Comments and conclusions

In the Kibbutz, children are reared by a nurse, live communally, and see their parents only for a short time during the day (usually in the evening). Thus these hypotheses follow naturally from such a regime. The fact that they were supported fits psychoanalytic theory but can hardly be regarded as very powerful evidence. It means little other than that the less you see your parents, the smaller will be the emotional attachments to them. This study is useful only in the context of other studies of the Oedipus complex, where it may be used as one small piece of evidence.

The work of Lasky and Berger (1959)

Lasky and Berger administered the Blacky Pictures to patients before and after genito-urinary surgery in a study relevant to the castration complex:

Sample. Thirty adult male patients. Age range 20-67 years, median 48.
Tests. The Blacky Pictures were given seven days before the operation and six days after it.
Results. Phi-coefficients of correlation were worked out for all the dimensions of the Blacky Pictures. The most affected dimensions (the smallest correlations) were: *masturbation guilt*, *castration anxiety* and *narcissistic love object*. These findings certainly support the hypothesis that genito-urinary surgery would produce castration anxiety.

Comments and conclusions

Although these results support Freudian theory, it must be remembered that the test-retest reliability of the Blacky Pictures is not high (Charen, 1956). Nevertheless the fact that the two dimensions most affected were those related to castration discounts to some extent the effects of low test reliability. It must be concluded, therefore, that this study supports the concept of castration anxiety.

The work of Hammer (1953)

Relevant to the study of Lasky and Berger (1959) is the study of symbolism among a sample of prisoners who had been eugenically sterilized (for sexual offences). These sterilized prisoners showed evidence of castration fear in their drawings on the House, Tree, Person Test. Chimneys were affected as were the branches of trees. Similar changes were not noted among a control group of surgical patients. Thus, in a manner similar to that of Lasky and Berger, Hammer also provides evidence for the existence of the castration complex.

The work of Sarnoff and Corwin (1959)

This was an experimental study of castration anxiety and the fear of death:

Rationale. Repressed castration anxiety is likely to be manifested as a fear of death. It therefore follows that individuals with severe castration anxiety should have a very considerable fear of death. Subjects with severe castration anxiety should show a greater fear of death than subjects with low castration anxiety after this castration anxiety has been aroused.

The following abbreviations will be used to describe this experiment since otherwise, as can be seen from the rationale, the repetitious language becomes too cumbersome:

HCA	High castration anxiety	MS	Morality scale
LCA	Low castration anxiety	HSA	High sexual arousal
FDS	Fear of death scale	LSA	Low sexual arousal

Hypothesis. HCA subjects will show greater fear of death than LCA subjects after being exposed to sexually arousing stimuli.

Design. A before and after design was used. There were two treatments (levels of sexual arousal) and two levels of castration anxiety, which was measured in a pre-experimental session.

Subjects. Fifty-six male undergraduates.

Method

A. *The pre-tests*
  1. Five item FDS
  2. Five item MS
  3. Blacky Castration Picture used to test for castration anxiety. On this variable, thirty-six LCA and twenty HCA subjects were selected.

B. A four week time-gap was allowed.

C. *The experimental sessions.* Two experimental conditions to which half of each group went. HSA: four pictures of nude women to which subjects wrote aesthetic responses. LSA: four pictures of fully-clothed models. Again aesthetic responses were recorded.

D. *The post-tests.* MS scale and FDS scale were again completed. In addition a questionnaire to assess degree of sexual stimulation was filled in.

Results

1. The HSA group were, in fact, more sexually aroused by the nude women than were the other group by the models.
2. HCA group showed a significantly greater increase in score on the FDS in the HSA condition than did the LCA group.
3. In the LSA condition there were no significant differences.
4. To test whether it was guilt causing the shift of scores, high and low subjects on the morality scale were compared on the FDS under both conditions. The difference was not significant.

Finding 2, that under conditions of high sexual arousal the fear of death score of those with great castration anxiety increased, while under low conditions it did not, is very strong support for the Freudian concept of the castration complex.

Comments and conclusions

As should be clear, this is an elegant and clever design. Any failures in the tests used would spoil the results so that it can hardly be argued that the tests were invalid. If they were, these results are even more remarkable. The fact that those with strong castration anxiety should, when sexually aroused, show a greater fear of death than those low on castration anxiety, can only be understood in terms of the whole Freudian concept of the castration complex. Finding 4 makes it clear that the results were not due to feelings of guilt.

Unlike correlational studies where the correlations can be due to errors common to both tests, e.g. response sets, and where the validity of the test must be proven beyond doubt, experimental studies, where conditions are manipulated according to a rationale and mean scores predicted, are more difficult to fault. They provide a far more rigorous test of any theory than do non-experimental studies, provided that, as in this case, the experimental rationale is relevant to the theory.

It must be noted that the castration complex is different from castration anxiety. It embraces much more: for example, it is related to fear of death and other symbolic injuries. Castration anxiety on the other hand is more specific, as in the study by Lasky and Berger (1959), where it refers simply to fear of genital injury. The complex refers to all the associated repressed material. This study by Sarnoff and Corwin provides strong support for the Freudian concept of the castration complex.

The work of Stephens

Stephens (1961) carried out a cross-cultural study of menstrual taboo using the method of Whiting and Child (1953). This study is also to be found in Stephens (1962).

Rationale. In *Totem and taboo* Freud (1913) argued that in the taboos, myths, superstitions and ceremonies of primitive societies could be seen phenomena which in our society are usually deeply hidden. For example, totemism was considered to be an institutionalized neurotic symptom caused by the Oedipus complex.

Hypothesis of the present study. The extensiveness of menstrual taboos in any society is determined by the intensity of castration anxiety felt by the men of that society.

Measures. Stephens (1961) argued that because there was no measure of castration anxiety *per se,* it would be necessary to measure the *antecedents* of castration anxiety. These are, according to Freudian theory, child-rearing procedures, especially those connected with the Oedipal situation. In addition he used three measures of the hypothesized *consequences* of castration anxiety–in this case, castration-like incidents in folk lore. Finally a menstrual taboo scale was used. Thus fifteen measures were used in all:

Ten measures of antecendent child-rearing procedures
The total scores derived from these measures
Three measures of the consequences of castration-like anxiety
One menstrual taboo scale.

Data on which measures were based. The anthropological reports of 72 primitive societies, as used by Whiting and Child (1953). Ratings were made for each of the variables.

Results. These are best set out in a table.

Child-rearing antecedent	*Relation to taboo scale in expected direction*	*Probability of chance occurrence*
1. Diffusion of nurturance	Yes	0.27
2. Post-partum sex taboo	Yes	0.01
3. Severity of masturbation punishment	Yes	0.01
4. Overall severity of sex training	Yes	0.05
5. Severity of aggression training	Yes	0.18
6. General pressure for obedience	No	0.65
7. Severity of punishment for disobedience	Yes	0.19
8. Strictness of father's obedience commands	Yes	0.20
9. Whether or not father is main disciplinarian	Yes	0.02
10. Importance of physical punishment as discipline	Yes	0.07
11. Total of variables 1 to 10	Yes	0.000001

The folk-lore incidents were not related to the menstrual taboo scale.

Comments and conclusions

It is clear from the results that the hypothesis is very strongly confirmed. This means, as Stephens (1961) points out, that menstrual taboos are affected by the castration complex, that there is a widespread phenomenon (72 societies) of castration anxiety and that it does originate in the Oedipal situation. It also implies that there is a widespread phenomenon of the Oedipus complex. All this is powerful evidence for the Freudian theory.

Careful scrutiny of the table of results, from the viewpoint of psychoanalytic theory, makes the evidence even more impressive. Thus the antecedent child-rearing procedures which would appear to be less relevant to castration anxiety are supportive of the hypothesis but to a lesser degree (e.g. variables 1 and 8) than those which are clearly most pertinent (e.g. variables 3 and 10). The enormously high statistical significance of the total score with the menstrual taboo scale is remarkable.

The fact that the folk-lore incidents did not also confirm psychoanalytic theory is curious but may only be due to the fact that other influences than castration anxiety affect the incidence of genital injury, severing and other kinds of physical injury in folk-tales. There must, for example, as in Homer, be historical antecedents as well as psychological ones.

This study by Stephens (1961) seems to constitute most impressive and powerful evidence for the Freudian concept of the castration complex.

The work of Whiting, Kluckhohn and Anthony (1958)

Whiting *et al.*, using the technique of rating societies, studied the function of male initiation ceremonies at puberty. In so doing they examined certain hypotheses concerning the Oedipus complex.

Rationale and hypothesis. The rationale for this study was derived from the anthropological speculations in *Totem and Taboo*. It was assumed that the ideal antecedent for the development of the Oedipus complex was a long *post-partum sexual taboo.* This, of course, enables the child to sleep with its mother. Similarly a short post-partum sexual taboo will tend to inhibit development of the Oedipus complex. Thus, by selecting societies with long post-partum sexual taboos and those with short taboos, the effects of the Oedipus complex should become apparent. The following hypothesis was formulated:

> Societies with a long post-partum sexual taboo (high on the Oedipus complex) will be more likely to have initiation rites (which are symptomatic of the Oedipus complex in psychoanalytic theory (Fenichel, 1945)) than will societies with a short taboo.

Sample. Fifty-six societies, representing 45 of the 60 culture areas designated by Murdock (1957).

Measures. The ethnographic literature relevant to these 56 societies was searched and ratings made.

Results. Eighty per cent of the tribes with long taboos had initiation ceremonies. The chi-square (sig. at 0.001 level) is hard to interpret in that through known historical connections the societies may not have been entirely independent cases.

Comments and conclusions

In the paper Whiting *et al.* concentrate on a careful examination of the background of various tribes in the sample, whose child-rearing procedures do not accord with their initiation rites, so that they are less interested than Stephens (1961) in testing the psychoanalytic theory. Nevertheless these results do support psychoanalytic theory concerning the Oedipus complex and its acting out in ceremony.

The work of Block and Ventur (1963)

This was an investigation of castration anxiety in a sample of amputees (symbolic castrates):

Test. The Blacky Pictures–the castration dimension.

Sample. Twenty amputees were compared with twenty surgical controls.
Results. The amputees were more disturbed on the castration dimension than the controls. Block and Ventur argue that this is support for the psychoanalytic theory of the castration complex and *ipso facto* for the validity of the Blacky Pictures.

Comments and conclusions

Although it is true that Freudian theory would predict an increase in castration anxiety among amputees, this study cannot be regarded as powerful support for the theory. This is because a simpler explanation of the findings is available. The castration card of the Blacky Pictures shows Blacky, blindfolded, with a knife suspended over his extended tail. It is not surprising that patients whose limbs have recently been cut off should be more upset at this picture than other surgical patients. This is not to impugn the Blacky Pictures. If, for example, homosexuals were higher scorers on this picture than normals, as Freudian theory predicts, a simple explanation would not be available.

This study, therefore, cannot rank high as supporting evidence for the Freudian theory of the castration complex. [. . .]

Appendix

The Blacky Pictures

The Blacky Pictures are a set of twelve cartoons portraying a family of dogs in situations which are peculiarly relevant to psychoanalytic theory. There are four dogs in this family–two parents, Blacky who can be of either sex (always the same as the subject taking the test) and Tippy, a sibling of Blacky. Subjects have to tell a little story about what they think is going on in each picture and how each of the characters is feeling. This spontaneous story is then scored for presence or absence of disturbance in the area concerned. In addition, after each story the subject is asked several multiple-choice questions about the cartoon. A further source of scoring is a spontaneous mention of a cartoon other than the one to which they are responding. A final task in this test is to sort the pictures into liked and disliked and choose the most liked and the most disliked. These two choices, according to the manual to the test, are symptomatic of disturbance in the relevant areas.

Variables measured and relevant cartoon	
1. Oral erotism:	Blacky being suckled.
2. Oral sadism:	Blacky tearing his mother's collar.
3. Anal expulsiveness:	Blacky defaecating between his parents' kennels.
4. Anal retentiveness:	As above, with different scoring.
5. Oedipal intensity:	Blacky watching his parents making love.
6. Masturbation guilt:	Blacky licking his genitals.
7. Castration anxiety (males) or penis envy (females):	Blacky watching Tippy, apparently about to have her tail cut off.
8. Positive identification:	Blacky playing with a toy dog.
9. Sibling rivalry:	Blacky watching parents fondle Tippy.
10. Guilt feelings:	Blacky having a vision of an angelic figure.
11. Positive ego ideal and	
12. Love object:	Variables 11 and 12 use two pictures differently scored for each sex. (*a*) vision of an ideal female dog (*b*) vision of an ideal male dog.

References

BERNDT, R. and BERNDT, C. (1951) *Sexual Behaviour in West Arnhemland.* New York: Viking Fund, Publications in Anthropology, 16.

BIDDLE, W. E. (1957) 'Investigation of the Oedipus fantasy by hypnosis', *American Journal of Hypnosis, 114,* 175.

BLOCK, W. E. and VENTUR, P. A. (1963) 'A study of the psychoanalytic concept of castration anxiety in symbolically castrated amputees', *Psychiat. Q., 37,* 518–26.

BLUM, G. S. (1949) 'A study of the psychoanalytic theory of psychosexual development', *Genet. Psychol. Monogr., 39,* 3–99.

CHAREN, J. (1956) 'Reliability of the Blacky Tests', *J. Consult. Psychol., 20,* 16.

DUSS, G. (1940) 'La méthode des fables en psychoanalyse', in DESPERT, J. L. (1946) 'Psychosomatic study of 50 stutterers', *Am. J. Orthopsychiat., 16,* 100–32.

FENICHEL, O. (1945) *The Psychoanalytic Theory of Neurosis.* New York: Norton.

FREUD, S. (1913) ed. STRACHEY, J. (1953) *The Standard Edition of the Complete Psychological Works of Sigmund Freud,* Vol. 13. *Totem and Taboo.* London: Hogarth Press and Institute of Psycho-analysis.

FRIEDMAN, S. M. (1950a) *An Empirical Study of Certain Psychoanalytic Propositions,* unpublished Ph.D. thesis, Western Reserve University.

FRIEDMAN, S. M. (1950b) 'An empirical study of the Oedipus complex', *Am. Psychol., 5,* 304.

FRIEDMAN, S. M . (1952) 'An empirical study of the castration and Oedipus complexes', in *Genet. Psychol. Monogr., 46,* 61–130.

HAMMER, E. F. (1953) 'An investigation of sexual symbolism: a study of HTPs of eugenically sterilized subjects', *Journal of Projective Techniques and Personality Assessment, 17*, 401–15.

HENRY, J. and HENRY, Z. (1944) 'Doll play of Pilaga Indian children', *Research Monographs of the American Orthopsychiatry Association, 4*, 133.

KUPPUSWAMY, B. (1949) 'Preliminary report of a statistical study of the Oedipus Complex', *Ind. J. Psychol., 24*, 289–335.

LASKY, J. L. and BERGER, L. (1959) 'Blacky Test scores before and after genito-urinary surgery', *Journal of Projective Techniques and Personality Assessment, 23*, 57–8.

LINDZEY, G. (ed.) (1954) *The Handbook of Social Psychology*. Reading, Massachusetts: Addison-Wesley.

McELROY, W. A. (1950) 'Methods of testing the Oedipus complex hypothesis', *Quarterly Bulletin of the British Psychological Society, 1*, 361–5.

MICHAL-SMITH, H., HAMMER, E. and SPITZ, H. (1951) 'Use of the Blacky Pictures with a child whose Oedipal desires are close to consciousness', *J. clin. Psychol., 7*, 280–2.

MURDOCK, G. P. (1957) 'World ethnographic sample', *American Anthropologist, 59*, 664–87.

POWDERMAKER, H. (1953) *Life in Lesu*. New York: Norton.

RABIN, A. I. (1958) 'The Israel Kibbutz (collective settlement) as a "laboratory" for testing psychodynamic hypotheses', *Psychol. Rec., 7*, 111–15.

ROHEIM, G. (1952) 'The anthropological evidence and the Oedipus complex', *Psychoanal. Q. 21*, 537–42.

SANOFF, I. and CORWIN, S. M. (1959) 'Castration anxiety and the fear of death', *J. Personality, 27*, 374–85.

SCHWARTZ, B. J. (1955) 'Measurement of castration anxiety and anxiety over the loss of love', *J. Personality, 24*, 204–19.

SCHWARTZ, B. J. (1956) 'An empirical test of two Freudian hypotheses concerning castration anxiety', *J. Personality, 24*, 318–27.

SEARS, R. R. (1943) *Survey of Objective Studies of Psychoanalytic Concepts*, New York: Social Science Research Council, Bulletin 51.

STEPHENS, W. N. (1961) 'A cross-cultural study of menstrual taboos', *Genet. Psychol. Monogr., 64*, 385–416.

STEPHENS, W. N.(1962) *The Oedipus Complex Hypothesis: Cross-cultural Evidence*. New York: Free Press of Glencoe.

TRILLES, R. P. H. (1912) 'Totémisme chez les Fan', *Bibliothèque Anthropos., 1*, 184–205.

WHITING, J. W. M. and CHILD, I. L. (1953) *Child Training and Personality*. New Haven: Yale University Press.

WHITING, J. W. M., KLUCKHOHN, H. and ANTHONY, A. (1958) 'The function of male initiation ceremonies at puberty', in MACCOBY, E. E., NEWCOMB, T. M. and HARTLEY, E. L. (eds.), *Readings in Social Psychology*, 3rd edn. New York: Holt, Rinehart and Winston.

WOODWORTH, R. S. and SCHLOSBERG, H. (1955) *Experimental Psychology*. New York: Holt, Rinehart and Winston.

6 The validity of Gottheil's Oral Trait Scale in Great Britain

P. Kline

Orality is an important concept in Freudian theory (Freud, 1905). Thus oral personality traits are held to be derived from fixation at the oral phase of development and certain psychosomatic diseases and symptoms are claimed to be related to oral fixation, notably peptic ulcer, alcoholism, and anorexia nervosa (Alexander, 1950). If this psychosexual theory were shown to be true, psychologists would have an important theoretical basis for the understanding of personality development. Hence the empirical study of orality is considered to be a valuable exercise.

Gottheil (1956a, 1956b) and Gottheil and Stone (1968) have developed and used a forty-item scale of oral traits. Its validity was attested (Gottheil, 1956b) by the fact that twenty *S*s [subjects] considered to be highly oral by a group of psychiatrists completed the scale almost entirely as hoped by the author. Twenty anally oriented *S*s presented dissimilar results on the same scale. These results were reversed on the anal scale. Psychiatrists are in considerable agreement that items of the oral scale do in fact reflect oral traits, i.e. face validity is high.

From this it appears, therefore, that the Gottheil oral scale fills a considerable gap in personality testing, especially for the empirical investigator of Freudian psychosexual theory. Oral scales are few in number and of unknown validity. Goldman-Eisler (1948) developed a rating scale while Barnes (1952) failed to isolate an oral factor. Krout and Krout (1954) with the KPPS [Krout Personal Preference Scale] and Grygier (1961) in his related Dynamic Personality Inventory (DPI) have produced oral scales but with little evidence of validity. Kline (1970) found little agreement between the oral scales of the DPI and Blum's (1949) Blacky Pictures. The problem of psychosexual measurement is fully discussed in Kline (1972), where it appears that the Gottheil scale is probably the best of the oral scales. The work of Gottheil and Stone (1968), who showed that oral traits were not related to mouth habits and that oral characters do not form a factor in P technique factor analysis, was held to be largely disconfirmatory of the Freudian theory of orality (Kline, 1972).

Method

In the course of some recent studies of the Freudian concept of orality, the Gottheil oral scale was administered to a large number of *S*s.

Source: *Journal of Personality Assessment,* 1973, *37,* 551–4.

Subjects

The Gottheil oral scale was administered to 241 students, 130 of these being postgraduate students of Education (mean age 21 years). The rest were a first year class in Human Development in the faculty of Social Studies (mean age 18 years). The sample was composed of 110 men and 131 women. This sample was chosen because of availability to the investigator and because both courses represent a wide cross-section of the student population.

The Gottheil oral scale was subjected to item analysis using the top and bottom 27 per cent method. Fan's tables (1952) were used which compute the percentage of *S*s putting the keyed response to each item and the correlation of each item with the total score.

Results

Table 1 sets out the results of this item analysis.

If the keyed response is put by more than 80 per cent of the sample or less than 20 per cent it is usual to reject an item simply because it is making too few discriminations. If, however, the correlation of this item with the total score were high it might be worth retaining. In personality testing it is usual to regard item total correlations of 0.3 as just about sufficient to make an item worthy of inclusion in a scale. Obviously the higher the correlation with the total score the better. With these rough criteria in mind we can now examine the results of the item analysis. Ten items correlate with the total score less than 0.2. This means that these items cannot be measuring the same factor as the rest of the scale. By the normal canons of test construction they would be eliminated from the scale. A further twelve items correlate with the total score below 0.3 and thus measure the common factor to only a slight extent. Thus twenty-two items of the forty appear to fail to form a scale. In addition two items (23 and 31) have so low a proportion putting the keyed response that they have little discriminating power. This means that only sixteen items in the scale are working as intended. It means, in effect, that with this large sample of British students the Gottheil oral scale scores could not possibly be valid since more than half the items were, as regards orality, compounded with error. Furthermore the fact that twenty-four items are failing means that the sixteen successful items cannot be expected to correlate very highly with the total. This is doubtless the explanation of the finding that the highest item total correlation is only 0.61 (item 32). Indeed the internal consistency of the scale, as measured by the Kuder-Richardson formula 20, is only 0.558.

It is instructive to examine the content of the successful items (ignoring statistical differences between them) to see to what extent it helps us understand the failure of the scale. A striking feature of the sixteen

Table 1 *Items on Gottheil Oral Scale*

Item	*P*	*r*
1. When you are troubled, do you talk it over with a friend?	0.65	0.54
2. Are you able to pout and act hurt when you want help or sympathy?	0.56	0.38
3. Are you inclined to be a jealous person?	0.49	0.25
4. Do you enjoy doing things on the spur of the moment, without planning or preparing?	0.81	0.27
5. Are you a talkative person?	0.55	0.40
6. Do you look on the bright side of things?	0.73	0.34
7. Do you study better with someone else than you do by yourself?	0.21	0.28
8. Are you influenced by people in authority?	0.67	0.28
9. Do you become upset because you do not have some of the things your friends have?	0.17	0.18
10. Do you enjoy strong competition?	0.56	0.09
11. Do you feel that you will get what you want out of life?	0.79	0.11
12. Do you feel that people should be forgiven if they apologise?	0.91	0.22
13. Do you do better when you are the 'underdog' than you do when you have friends on your side encouraging you?	0.40	0.18
14. Do you become upset when your friends forget your birthday?	0.43	0.18
15. When you like someone, can you tell him so?	0.61	0.42
16. Do you feel the world is a happy place?	0.46	0.37
17. In a new town, would you prefer to explore it by yourself rather than have a friend show you around?	0.35	0.39
18. Do you like being the leader of a group?	0.54	0.16
19. If you come across a word you do not understand, do you look it up in a dictionary rather than ask someone?	0.48	0.46
20. Do you keep your emotions locked up inside you?	0.39	0.48
21. If there were no difference in salary, would you rather be a bookkeeper than a salesman?	0.53	0.26
22. Do you feel that sympathy hurts a person more than it helps him?	0.77	0.23
23. Can you enjoy an evening alone by yourself?	0.14	0.32
24. Do you make an effort to get on your superior's 'good side'?	0.45	0.47
25. Do you do things without thinking and later regret what you have done?	0.71	0.30
26. Can you express yourself better in writing than in speaking?	0.53	0.42
27. Do you try to take on responsibility?	0.45	0.11
28. Are you more sentimental than most people?	0.43	0.30
29. Do you have a weak personality?	0.22	0.15
30. Do you feel that you can learn more from reading than you can from talking with people?	0.79	0.27
31. Do you make major decisions on your own?	0.17	0.41
32. In a restaurant, do you buy more than you can eat?	0.16	0.27
33. Assuming the same income, would you rather own your own business than work for a large and respected company?	0.25	0.28
34. Would you want to be alone when you receive bad news?	0.38	0.29
35. Do you continue to follow an idea of your own even when most others disagree?	0.21	0.21
36. Do you like to have a friend go with you when you buy your clothes?	0.35	0.35
37. When you are sad, do you try to find someone to talk you out of it?	0.60	0.61
38. Do you find it hard to ask people for advice?	0.68	0.51
39. Do you think people can get ahead without 'playing politics'?	0.26	0.08
40. Do you try to make friends with people who are more successful than you?	0.18	0.14

P = proportion putting keyed response

r = correlation of each item with total score successful items is that a high proportion of them is concerned with what might be called talkative sociability or communicativeness (Items 1, 5, 15, 17, 19, 20, 26, 36, 37, 38), ten items in all. An ability to verbalize feelings is evident. This quality is of course central to extraversion as conceived by Eysenck in the Eysenck Personality Inventory (EPI) and item 1 is virtually a paraphrase of an EPI extraversion item. These ten successful items, therefore, as regards content would appear to measure social extraversion. Looked at in this light, the six other successful items fit this picture, especially 25, impulsivity, being largely concerned with cheerful optimism.

Discussion

From this study of the item analysis of the Gottheil oral scale several clear points emerge:

1. With a British student sample, the items fail to form a scale, and the test must be invalid. This implies that it cannot be used with any British sample. Since, to some extent, American tests can be used in Britain, it casts some doubt on the American results.
2. A study of the successful items reveals that the common factor underlying some of the items is extraversion. Thus the Gottheil oral scores appear to be crude extraversion scores; crude because compounded with error.
3. It follows from this that the earlier studies with this scale must be viewed with great caution and that the results of Gottheil and Stone (1968) cannot be held to either confirm or refute the Freudian theory. The importance of these findings is considerable in the light of the need for firm evidence about Freudian psychosexual theory.
4. Clearly the validity of the oral scale needs further examination. More items should be written, the scale should be tested afresh and the common factor measured should be identified not by item content but by its correlations with other scales and criteria.

References

ALEXANDER, F. (1950) *Psychosomatic medicine.* New York: Norton.

BARNES, C. A. (1952) A statistical study of the Freudian theory of psychosexual development, *Genet. Psychol. Monogr., 45,* 109–74.

BLUM, G. S. (1949) 'A study of the psychoanalytic theory of psychosexual development', *Genet. Psychol. Monogr., 39,* 3–99.

FAN, C. T. (1952) *Item Analysis Table.* Princeton, New Jersey: Educational Testing Service.

FREUD, S. (1905) 'Three essays on sexuality'. In *Complete works of Sigmund Freud. Vol. 7.* London: Hogarth Press.

GOLDMAN-EISLER, F. (1948) Breast-feeding and character formation. *J. Personality, 17,* 83–103.

GOTTHEIL, E. (1956a) 'Conceptions of orality and anality', *J. nerv. ment. Dis., 141,* 155–60.

GOTTHEIL, E. (1956b) 'An empirical analysis of orality and anality', *J. nerv. ment. Dis., 141,* 308–17.

GOTTHEIL, E. and STONE, G. C. (1968) 'Factor-analytic study of orality and anality', *J. nerv. ment. Dis., 146,* 1–17.

GRYGIER, T. G. (1961) *The Dynamic Personality Inventory,* London: National Foundation for Educational Research.

KLINE, P. (1970) 'A projective and psychometric study of the oral character'. *Proceedings of the 7th International Conference of the Society for Rorschach and Projective Techniques.* Vienna: Hans Huber.

KLINE, P. (1972) *Fact and Fantasy in Freudian Theory.* London: Methuen.

KROUT, M. H. and KROUT, T. J. (1954) 'Measuring personality in developmental terms', *Genet. Psychol. Monogr., 50,* 289–335.

7 Personality testing and the school child

R. B. Cattell

Personality and school achievement

It is a commonplace among experienced observers in almost any field of endeavour that personality and ability together decide the outcome, their relative importance varying according to the form of achievement involved. It should be understood that we include here, under personality, also motivation and values.

Consequently the educational psychologist has hitherto taken a peculiar position, for in fact he has devoted his efforts to predicting achievement from abilities alone. For example, when schools in Britain select pupils for the eleven plus scholarship selection, or for entry to the university, they do so by an academic attainment examination and by an intelligence test. And when a psychologist in a child guidance clinic looks for the causes of backwardness in a child referred for very poor school performance, his examination begins–and all too frequently ends–with an intelligence test. Until a few years ago this limitation was necessitated by the absence of valid tests of personality and motivation. As measures are now available covering much of the personality and motivation area, then the practising psychologist should not be denying the new resources held out to him, but should revise his ideas and rise to the challenge of new powers–and, of course, new problems!

The first information on the extent to which personality and motivation measures affect and predict school performance has come from studies on normal children by Pierson, Sweney, and Sealy in America, Connor and Radcliffe in Australia, and Butcher and Warburton in Britain. There have also been special studies on personality influence in backward children by Wright, O'Halloran, and others. Immediately, it has become apparent that personality factors do indeed contribute just about as much as abilities themselves to school achievement. Many a schoolteacher would have been ready to bet on this conclusion without the evidence of tests, but the test experiment brought objective confirmation of the magnitude of these influences and also on the directions in which temperament and dynamic traits act.

When the High School Personality Questionnaire, which measures fourteen distinct factors, and requires no more than an hour's testing time, has been given to groups of some 300 boys and girls, of around twelve years of age, or the 16PF [Sixteen Personality Factor test] to students of

---

Source: CATTELL, R. B. (1965) *The Scientific Analysis of Personality*. Harmondsworth: Penguin, 309–25.

eighteen, the 'specification equation' obtained averages as follows:

$$\begin{aligned} A_c = &+ 0.2A + 0.5B + 0.3C - 0.2D - 0.2F + 0.4G + 0.2H \\ &- 0.2I - 0.2M + 0.3Q_2 + 0.3Q_3 - 0.1Q_4 \end{aligned}$$

This formulation states that a child's achievement score in English and arithmetic combined (A_c) is best estimated by weighting his affectothymia temperament score (A) positive, his intelligence score (B) positive, his ego strength (C) score positive, his excitability (D) negative, and so on, by the decimal fractions indicated, and then adding the lot (in standard scores). Incidentally, different research articles have reported slightly different weights, accorting to type of school, age of subjects, country, etc., and as discussed below, we should, psychologically, expect such variation. The reader may also be reminded that these weights are approximately the same as the *correlations* of the given personality factors with the 'criterion', the child's score in the basic subjects–English and arithmetic. They show how much the given personality source trait concerned helps or hinders achievement and when we look at their natures the results make good sense.

For example, other things being equal (and this is what we mean when we speak of a weight for a given factor) a child who is more outgoing, adaptable and warmly related to the teacher (all aspects of A factor) will learn faster (+0.2). The child who is more emotionally balanced and less easily upset (C factor, ego strength) will also learn more easily (+0.3). The individual of greater conscientiousness (G factor) will make more progress (+0.4) for the same intelligence, and so on.

Nevertheless, not all these personality connexions discovered by research are so self-evident. For instance, the more dominant child (E factor), according to these measurements, learns more slowly than the more submissive. Why should this be? Some perspective is given on this by the finding that at a much later stage, with post-graduate university students and scientific researchers, the reverse is true. The more dominant turn out to be more creative and to achieve more in the field of research. As pointed out in discussing the nature of dominance, the docility and imitativeness which make a good examination-passer are not what make a good independent, critical thinker. Some may regard this finding as obliquely exposing something wrong with our school system. This is not necessarily true: perhaps class teaching is only possible with some degree of docility. But it raises a deeper question, discussed below, concerning whether a scholarship should select for good school or good *post*-school performance.

The weights on various personality factors will also change with the school subject concerned. They are, in fact, somewhat different for performance in English and performance in arithmetic, the subjects which we have averaged above to give a general school performance. In this respect it is surprising to find that introverts (to be exact, sizothymes–$A-$–and desurgents–$F-$) tend to have better vocabularies and grammatical skill than extroverts, who actually do more talking. A simple possible explanation is that introverts read more, and the people they meet in reading have better vocabularies and grammatical styles than the people they would meet in everyday life. But it is also possible that they love words more because they live more in a world of ideas than with concrete things.

The manner and degree in which a personality factor aids the expression of ability in producing achievement also varies with circumstances and age. For example, the big temperament dimension, affectothymia-*vs*-sizothymia, or *A* factor, actually changes its direction of influence somewhere between twelve years of age and the university student period. In the former, affectothymia correlates positively with classroom achievement, whereas among students all of equal intelligence the sizothymes get better grades. The loading or weight changes from +0.3 to −0.4! Presumably we are seeing here a change from social reinforcement of effort in the classroom, well responded to by the sociable affectothyme, to more individual, internal direction at the university level where the sizothyme enjoys his autonomy. In the university, also, the largest positive loading shifts to Q_3, i.e. the person with a well-integrated self-sentiment gains the advantage.

Whereas it is very rare for any known *ability* to load an achievement performance negatively, personality factors are frequently good for one kind of achievement and bad for another. Among various abilities the worst that can be said of, for example, musical ability is that it is unimportant in contributing to one's drawing ability. It does not *interfere.* (Having a higher spatial thinking ability, as Thurstone showed, actually goes with slightly better than average verbal ability.) The ambivalent action of personality factors, on the other hand, is shown by *A* and *E* above, and also by the sharp contrast between the negative loading of surgency, *F*, in examination success and its substantial positive association (perhaps due to liveliness, talkativeness, and limelighting) with effectiveness in small groups and election to positions of leadership. Again, desurgency–*F* (−)–as Drevdahl has shown, is characteristic of original thinkers. This opposition between popularity for elections, and original, creative thinking may well explain why democratic leaders so seldom lead! There are, however, countless other instances where personality factor weights flip over from positive to negative with different kinds of achievement. This is the personality analyst's expression of the fact that 'it takes all kinds to make a world', though, to be realistic, the uses for some kinds of personality endowment are hard to discover!

The effects of motivation and values on achievement

It is commonly assumed in the exhortation 'try your hardest' that the relation of motivation to achievement is a simple one. It is certainly a powerful one, but not always simple. For example, performance on intelligence tests fails to improve beyond a quite moderate level of effort and may decrease. As Sarason has shown, when people are at a high anxiety level greater effort may actually lead to reduced performance, notably in committing to memory.

Our evidence here is largely on various ergic tension levels and on sentiment structure development in relation to school achievement. To understand the answers it is necessary to introduce one more statistical notion, namely, that the percentage of the variance (variability) in a certain criterion accounted for by some personality (or other) measure which predicts it is the *square* of their correlation, *not* just the coefficient itself. For the present this may need to be taken on faith, but obviously it is a useful notion when we want to account for the fractions out of a total performance which are associated with particular sources.

For example, the correlation typically found between an intelligence test and school achievement is around 0.5 (0.55 in a recent survey in Scotland, but perhaps Scots are a little more careful not to waste any intelligence!). This means that $(0.5)^2 \times 100$, i.e. 25 per cent of the range observed in achievement is associated with the intelligence variability in the particular population. Or, looked at another way, if we could instantly eliminate differences of intelligence–or take a lot of people all chosen to be just at the same intelligence level–the variance in school performance would still be appreciable, namely, 75 per cent of what it now is.

This kind of statistical calculation is very useful in giving us an idea of the relative importance of various causes–and also in showing us how much we still do not know. For the various contributions should add up to 100 per cent and no more. Now we are approaching the question of the relative importance of abilities, personality traits and motivation factors. Many experiments have converged on reliable estimates for the abilities. If we use several primary abilities, as in Thurstone's tests, we may predict a shade more than from intelligence alone, but the figure usually stands at 25–30 per cent.

An estimate for personality factors may be obtained approximately from squaring the correlations in the equation above. (Actually, because of rounding and some correlations among factors the literal value is too high, and the best estimate would be roughly 40 per cent.) The experimental evidence on motivation measures, with MAT (The Motivation Analysis Test) is too recent for any precise conclusion to be drawn until checked, but it looks as if about 20 per cent will be found associated with motivation. Allowing for slight overlaps among the measures from these three *essentially* independent 'modalities' of psychological measurement, a shrewd estimate might be that the three batteries will account for 60–80 per cent of the variance in achievement.

To some unfamiliar with general psychological measurement this may not seem a very high efficiency of total prediction. But actually we should be surprised and suspicious if it were much higher. For this tells us how much we can predict from *qualities within the student*, and, as every student well knows, something depends on relatively accidental circumstances outside himself, e.g. an illness, a good tip on what is likely in the exam, a lost notebook, the conditions for working at home, the attitudes of his parents and his peers, and so on.

As to the particular dynamic factors which seem systematically to associate themselves with achievement the work of Connor, Horn, Pierson, Sealy, Sweney, and others already converges, on their varied student groups, to a decided unanimity of conclusion. This is summarized—necessarily roughly because of the absence of common units—in Figure 1.

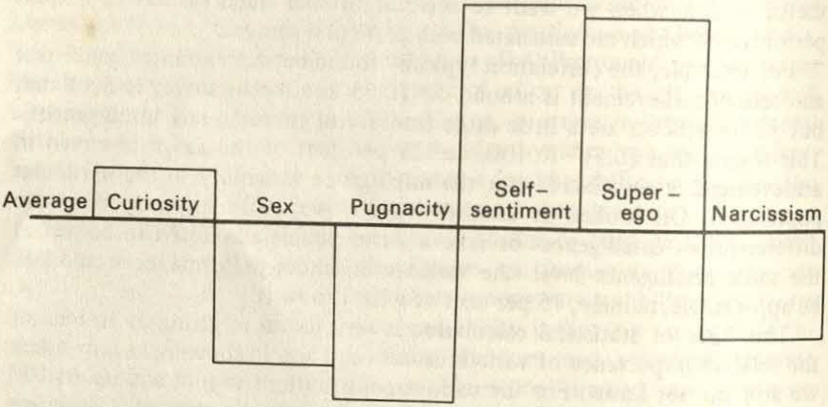


Fig. 1 The dynamics of school achievement. This profile of the high achiever compared with the low achiever shows on what dynamic structures the former tends to excel and to be below average.

Some of these differences fit other sources of evidence. For example, since the self-sentiment and the super-ego are already known to load achievement when measured on the 16PF questionnaire, it is not surprising that they should do so also when measured by these objective motivation strength measurement devices. Nor is it surprising that the general strength of curiosity should advance learning.

The quite large negative roles found for ergic tension levels on sex, pugnacity, and narcissism, however, are striking and will lead to some new developments in our thinking. A simple type of explanation of the effect of sex tension level might be that the student experiencing higher drive is more distracted from his studies. A more sophisticated and perhaps more likely one is tied up with Freud's generalization that culture is sustained by sublimated sex drive. Possibly the individuals who have achieved more of such sublimation, and thus reduced the tension for direct sex expression, do better because of these greater resources for cultural attainment.

On the other hand, the higher pugnacity is probably better explained as a consequence than as a cause of poorer achievement. Pugnacity (which is, incidentally, a demonstrably quite different factor from the self-assertiveness erg) is the consequence of a frustration situation. And at the stage of life when school work occupies so large a sector of the life circle, in a competitive school system, what could generate more pugnacity and sadistic tension? Although MAT measurements do not yet exist for pre-schizophrenics, one sometimes sees a high build-up of pugnacity in the late adolescent who repeatedly fails in occupational try-outs and moves towards a schizophrenic intransigence.

Alternative explanations as cause and as effect have also been offered in response to this new finding that narcissism has a quite powerful negative relation to achievement. The narcissistic drive is a general sensuality, with a sexual quality, but directed more to the individual himself, much as psychoanalytic clinical observations have always described it. However, the factor pattern shows a somewhat broader quality, with self-indulgence, avoidance of work, selfishness, rejection of moral demands, and much which, in a group, would be the very antithesis of high morale. Some clinical explanations have considered this to be essentially a 'regression', and doubtless certain early massive discouragements might contribute to it. But it seems far more probable that the poor achievement is a consequence of this pervasive self-love.

Some finer issues in motivation and achievement

The two preceding sections have tried to show the results of recent advances in ability, personality, and motivation testing, indicating what each area contributes when the other is held constant. For example, the contributions from dynamic tension levels assume intelligence to be the same when the achievement differences are affected by motivation differences. It has been shown that the three modalities are very roughly equal in their contribution to achievement and that they account for about two-thirds to three-quarters of the 'explanation' (variance) of individual differences in performance. Although, as Tennyson reminded us, 'science moves but slowly, slowly, creeping on from point to point', and these findings of the present pioneer decade will need to be confirmed in diverse situations, the potential advance of educational practice from what could be done with ability tests alone is almost staggering.

Before trying to focus some conclusions for practice, and some social implications, let us consider a little more the full scientific complexity of these findings. First, how general are they? Actually, the results of the American investigators above have already received reasonable confirmation by Radcliffe in Australia and by the lively work of Butcher and Warburton. Much the same personality and motivation factors are active, though there are some slight but interesting differences of emphasis too. Secondly, how do mechanisms vary with different ages of students and different type of educational system? Thirdly, how do the three kinds of trait interact?

As regards the latter, it is reasonably sure that certain abilities and general personality factors determine interests rather than vice versa. For example, how can anyone get very interested in the amino acid structure of DNA or excited about Bessel functions if he does not in the first place have the *B* factor (intelligence) level necessary to *see* the relations? [. . .]

Regarding the second question above, the predictive weights required for various personality and dynamic factors will obviously change, statistically, with local social characteristics and school selection. The 'importance' of traits will vary.

University students, for example, have already been selected for intelligence, so one would no longer expect differences of intelligence to be so important in predicting the differences in achievement among them. In fact, one would anticipate that personality differences would now climb to the paramount position as the decider of grade performance–and recent results support this. Secondly, one should note that the weights will also vary with the situation in which the learning is carried out. For most students the change from the ordinary school, where a teacher treats them with the solicitude of a good hen for a brood of children, to the university, where they are (more in European than American universities) expected to be their own managers, is one which takes some adjustment. New qualities of personality come into action. We have already seen that the loading on *A* factor, affectothymia-sizothymia, actually seems to reverse its sign. This says that students who respond strongly to the social milieu do well in school, but that the detached and self-sufficient do better at the university.

In particular we can expect a more complex set of laws to develop regarding prediction, especially from motivation measures. In the first place these measures will not be so relatively constant from month to month as will ability and general personality source traits. A drive tension level in particular may change with the general stimulation level of the environment and the degree of intercurrent satisfaction. We need to know more about how big these changes are and how changes of achievement in turn affect them. Again, some curious differences have appeared between the measurement of a dynamic structure in its unintegrated and its integrated motivation components. For example, the self-assertive erg in its integrated manifestations seems to contribute positively to achievement, whereas the scores on its unintegrated component battery are definitely negatively related to achievement. We have much to learn about these differences of meaning among the distinct motivation components, and when these insights are gained our control and prediction should reach new levels.

The relation of personality and motivation scores to performance throw light not only on the individual, but on the school system. Differences can be seen between the personality and motivation structures which command success in one type of school and in another. In the last two or three generations there has been more concentration in the art of the teacher on attaching new ideas to existing interests of the child instead of driving him, like a donkey, with blows from behind. Different personality traits will obviously contribute to success under two such different systems. What is perhaps constructively of more importance is that the emergence of reliable, objective measures of dynamic strength will enable the educator to get reliable information about the strengths of sentiments on which he can hope to build at various ages and with various types of student.

The facts already cited above–that sizothymia is negatively related to performance in grade school and positively in the university; that docility (low *E*) favours examination-passing but is a defect when research work and creative art are the goals, etc.–may suffice to show, however, how measurement in schools can help to throw light on the teacher and the accepted institutions as well as on the individual.

Finally, research is beginning to throw light on the meaning of these personality and motivation factors in achievement considered in a broader context than the school alone. If the selective school is preparation for something *beyond* the school, then we should be selecting for it by the qualities which give success later, rather than those which give success only in its examinations. In these later successes there is reason to believe that the balance of weighting tips over and that dynamic traits, instead of being third in order of their contribution to achievement, as in the school, become paramount.

Some practical problems in school psychological services

Let us now consider the implication of the above findings for an efficient school psychological service. It has already been pointed out that such a service must be many-sided. It must think of at least four aspects: (1) keeping records of achievement and making predictions and scholarship selections; (2) understanding the emotional and social adjustment of the individual child, at recurrent 'problem' occasions, as in Child Guidance Clinic work; (3) giving vocational guidance at the close of the school period; and (4) understanding what the classroom institutions and teaching methods are demanding of the pupils.

Many school systems in advanced countries supply at least the first three of these. But all too often they do so through different persons and offices, using testing devices which do not interlock and records which are not easily mutually available. In view of the scientific advances which have oriented testing to organic realities in the individual, psychological testing, like medical testing, has become integrated, and the administrative segregation of functions like child guidance treatment of problem children, special school treatment of backward children, scholarship examination, a guidance teacher's handling of vocational guidance, etc., is a wasteful disregard of modern possibilities. A school psychological centre should handle all of these with one or two fully qualified psychologists (or more in a city centre). Records of the total personality and ability would then be preserved from the beginning at the central school psychologist's office. For example, the psychologist handling clinical cases would not be left to find, the first time that a child is referred for help, that no earlier data whatever exists on his former personality and interest development, nor would the examination of scholarship potential be a mechanical procedure divorced from knowledge of the emotional make-up and interest of the particular child.

However, the more effective and far-reaching the psychological care of the child becomes, the more the psychological scientist is going to have to come out of his laboratory backroom and justify and explain his procedures to the public. As pointed out earlier in this chapter there have been minor revolts against his procedures, largely rationalizing an unrealistic, narcissistic wish that one should not be measured against environmental demands, but partly on the misunderstanding that a psychologist commands when in fact he merely advises. However, as present-day examples of how involved these issues may get in popular superstition and political slogans one may instance the burning of personality test papers in Texas and the decision of a socialist London County Council to abolish the scholarship examination's recognition of individual differences in ability, on 'the principle than no London child must be labelled "suitable" or "unsuitable" for any particular form of education' (*Evening Standard*, 1 July 1963). The danger of scientifically uneducated politicians, writers, and journalists propagandizing public opinion in a world where the understanding of man and society has also become an intricate science, with many gifts to bestow, is nowhere better illustrated than in the schools. To show that science can be misunderstood from any political angle, a conservative newspaper of the same city (*Daily Telegraph*, 2 July 1963) said in an editorial on scholarship selection: 'Some of the greatest men of our own time–Sir Winston Churchill is a primary example–showed singularly little ability . . . in boyhood.' These varying types of development are well understood by psychologists. Churchill, Darwin, and others would have been recognized by intelligence tests whereas they were not recognized by ordinary school examinations. Churchill is clearly an example of, among other things, the lawful relation on page 79 above, where high dominance factor, *E*, detracts from school performance but contributes to original performance in later life.

These arguments are against poor, unimaginative, and bureaucratic use of psychological tests, not against more effective use. But it is ironical that the writers think the solution is to go back to no selection at all–the bright and the backward going at the same frustrating lock-step in the same classroom–or to selection by interviews, essays, and teachers' all-too-human personal prejudices. This is done in the city where Sir Cyril Burt, around 1910, pioneered with the most up-to-date psychological testing programme then in existence and published, in 1921, results demonstrating solutions which have become classical in mental testing, and which fully answer the confusion voiced in 1963.

Misunderstandings, by parents or politicians, of what tests aim and do not aim to do, and what degrees of accuracy they can and cannot attain, are, nevertheless, being increasingly eradicated. But wider confidence in testing can come only if certain real technical shortcomings of psychological measurement are faced. For example, as tests are recognized to affect positions, salaries, and practical affairs, there will be increasing efforts to 'beat' them, by coaching, faking, etc. That coaching for intelligence tests can be met by coaching everyone, with effective reduction of test sophistication differences, has been shown by Eysenck. The issue of faking in personality tests is a more complex question, which needs some essential comments.

Psychologists have recognized and experimented upon the tendency in questionnaires to 'fake good' and give 'socially desirable' responses. When subjects respond as truthfully as possible, in an anonymous situation, and then do the same test again where they are limelighted and the results have bearing on their future, the latter situation turns out to produce an aberration which is not along a single 'social desirability' dimension but along several dimensions, turning upon the 'role' which the subject chooses to play.

There are three ways of handling this 'motivational distortion' influence. First, by hiding in the test certain items aimed to evaluate the degree of 'faking'. Thus the MMPI [Minnesota Multiphasic Personality Inventory] has a 'lie scale' and a 'faking scale', etc., and the 16 PF has an MD (motivational distortion) scale. These may assume, as an approximation, a single distortion shift, and correct each personality factor score by its amount. However, in the broader sense psychologists have to recognize two major possibilities in test taking: motivational distortion and sabotage. The latter is rare, but we have to recognize that in every society there are one or two who take perverse pleasure in wrecking. For example, a non-cooperative person might check every alternative answer, or try to answer randomly. By certain probabilitity tests, applied by computer, such departures from any natural structure can almost infallibly be recognized. Nothing can at present be extracted from such a response sheet, but at least it can be set aside as sabotaged. The motivational distortion problem, on the other hand, is comparatively small and looks as if it might be solved in a decade.

Secondly, one may seek items which correlate with the required personality factor but do not obviously look as if they should, i.e. have no face validity. This is a promising line and has been followed in the High School Personality Questionnaire. [. . .] Lastly, one can shift from questionnaires to objective tests. Because of the greater demands on technical psychological skill, and the increase of the testing time demands by a ratio of 4 or 5, this valuable step forward is only being made slowly.

A second major respect in which testing must put its house in order as it comes increasingly into practical life is in demonstrating more regard for what should be the real criterion. For example, in the stress of war conditions military psychologists had to accept the infantry officer schools' grades as the 'criterion' of a good officer, there being no opportunity to record subsequent behaviour in the field. The ability and personality tests used succeeded in so cutting down the rejection rate among those taken into training that millions of man hours and dollars were demonstrably saved. But still the criterion was an artificial one, and it was not until the Korean War that records of actual battle behaviour were meticulously obtained and used as the true criterion against which any test had to stand up in prediction. Less dramatically, but equally realistically, in peace the psychologist has to adjust his selection to the criterion of 'the long run' rather than some immediate performance.

Actually the tendency to depend on the cheap and immediate criterion has nothing to do with psychological testing. The psychologist develops his tests of personality structures and the prediction he gives is usually more limited by the unreliability of the criterion he is given than by his own tests. The criterion for scholarship selection, for example, should not be success in school but success after the school. Shrewd observers in education have long objected that pupils who are skilled examination-passers, and who may also do well in intelligence tests, do not always live up to the promise after school. They may be colourless, or lack enterprise, or fail to show a sturdy interest in hard work. Though success in particular occupations to some extent requires particular traits, there are certain traits which are necessary for successful use of one's abilities in almost any area, and it would not be unreasonable in awarding school scholarships to give weight to those traits important for the occupations served by such schools. Such weight would almost certainly raise personality factor scores to the same order of importance as intelligence tests. The ordinary school examination may actually weight these at present in the wrong direction, as in the above instance of higher score on dominance (*E* factor) being associated with greater creativity in art and scientific research but lower score in exams, whereas higher docility systematically contributes to good examination-passing.

Yet another social aspect of psychological testing arises from the extreme complexity and many years of necessary research in modern test production. The day of the 'patent medicine' test, made up by the individual by a stroke of alleged inspiration, is gone, as also is the test made up specially for the occasion by the psychologist, like the old-fashioned doctor going into his back room or garage to concoct his own pharmaceuticals. There is indeed considerable parallelism between the great pharmaceutical houses, developing in their research laboratories biochemical and other 'drugs' of a complexity beyond the individual pharmacist's scope, and psychological test publishers who supply the psychologist. The question of whether there should be a single state monopoly or a careful arrangement of adequate competition among several private enterprises is one which is likely to be answered too frequently by a political response. In America there are half a dozen major independent test research centres–the Psychological Corporation founded by McKeen Cattell; the Educational Testing Service, supported in part by universities and colleges; Science Research Associates, founded by Thurstone; IPAT, the Institute for Personality and Ability Testing, and certain general book publishers, e.g. the World Book Company, which maintain test development centres in which standardizations, relations to criteria, and improved scoring facilities are worked out. This variety has permitted an open-mindedness to new possibilities which a single bureaucratic 'authority', unassailable by competition, has usually not shown. At any rate, it is unquestionable that America has gone farthest in technical level of test development, in providing computer and other facilities, and in 'lending' its tests to other cultures.

One substantial advantage of the factored ability of personality tests, such as the Thurstone Primary Abilities Test, the 16 PF, the G–Z questionnaire, the O.-A. Batteries, etc., is that the criterion relationships, the occupational profiles, the clinical diagnostic weights, etc., which are built up at such cost of human effort, do not have to be thrown away when the test is scrapped or revised. With a 'patent' test, as with a patent medicine, all we know is that a certain amount predicts a certain result, and we have no idea through what functions it works. On the other hand, when we establish that so much general ability (Spearman's '*g*', our *B*) is required for a certain occupation, or that surgency (*F*), and parmia (*H*) are respectively positively and negatively correlated with prognosis in a mental disorder, this information preserves its value and meaning when, after perhaps ten years, some slight improvements are made in the method of test measurement of, say, surgency, parmia, or intelligence.

8 Primaries or second-order factors: a critical consideration of Cattell's 16PF battery

H. J. Eysenck

There are considerable similarities between the personality descriptions given by the factor-analysis based systems of Cattell, Guilford and Eysenck; these similarities, however, appear only in the higher-order factors called extraversion–introversion and neuroticism–stability by Eysenck, and exvia–invia and adjustment–anxiety by Cattell (Eysenck and Eysenck, 1969). While the factors extracted at this level from sets of questions contributed by these three authors are virtually identical, there is little agreement on primary (first-order) factors; furthermore, Eysenck was unable to replicate Cattell's or Guilford's factor on his samples (Eysenck and Eysenck, 1969). As far as Cattell's primaries are concerned, this failure seems common; Peterson (1960) in the USA and Grief (1970) in Germany similarly failed to provide any kind of replication. It could be argued, in the case of the Eysenck and Eysenck studies, that the method of rotation used was dissimilar to that used by Cattell; however, Promax (Hendrickson and White, 1966) was devised in our laboratories by two former collaborators of Cattell's with intent to capture the principles underlying his own methods, and a study carried out in Cattell's own laboratory has shown Promax to be equivalent to Cattell's own methods (personal communication).

Even if Cattell's factors were replicable, there would still be disagreement on the relative importance and value of primary factors and of second-order factors. Cattell is quite specific in his claims: 'The primary factors give one most information, and we would advocate higher strata contributors only as supplementary concepts. . . . It is a mistake, generally, to work at the secondary level only, for one certainly loses a lot of valuable information present initially at the primary level' (Cattell *et al.*, 1970). Eysenck's position is equally clear; he would maintain that second-order factors are far more meaningful psychologically (Eysenck, 1967a), and that little if any information is lost by disregarding the primaries in such personality studies as those reported by Cattell. It is not suggested that this would always and inevitably be so; in the field of intelligence testing Eysenck would still regard *g* as the most important single factor, but would certainly agree that half a dozen or so of primaries make a definite contribution to prediction in many cases (Eysenck, 1967b; Vernon, 1965; McNemar, 1964).

Such an argument should be amenable to factual settlement, and the present paper constitutes an attempt to provide some information which may be considered relevant. Cattell *et al.* (1970) have published a table of intercorrelations between the 16PF scale scores of 423 male and 535 female

---

Source: *British Journal of Social and Clinical Psychology*, 1972, *11*, 265–9.

college students, separately; they used the sum of scores on forms A and B for this purpose. Many of these are quite high, but for our purposes these raw correlations are not very useful as they are of course very much lower than the 'true' correlations between scales by virtue of the rather low reliabilities. Our argument will be that if the correlations between scales contributing to a particular factor (exvia, or anxiety) at the second-order level are at or near unity, then clearly the individual scales (primaries) make no contribution over and above that made by the second-order factor. To test this hypothesis we must correct the existing correlations for attenuation; without this correction the failure of the observed correlations to reach unity may be due entirely to random error rather than to factor-specific contributions. For this purpose, reliability coefficients are required for the scales used; fortunately correlations between scales A and B which Cattell *et al.* consider 'parallel forms' are given in their Table 5.3 for 6476 subjects, and have been used for our calculations. It would have been more suitable if these reliabilities had been calculated on the actual sample studied, but the requisite data are not available; this may introduce some minor errors into the computations. On the other hand, the reliabilities of the figures may be considerably improved because they are based on much larger numbers; it is possible that advantages and disadvantages balance out in this case. This correlation between parallel forms, Cattell calls 'equivalence coefficient'; as he points out, there are many different 'reliability' coefficients in the literature, all having quite different properties, and it is necessary to be quite specific about one's use of the term.[1]

Table 1 *Intercorrelations between Cattell's five 'anxiety' scales, corrected for attenuation*

		C(–)	L	O	$Q_3(-)$	Q_4
C(–)	(Low ego strength)	–	1.14	1.22	1.04	1.19
L	(Suspiciousness)	0.98	–	0.86	0.78	1.15
O	(Guilt proneness)	1.24	0.66	–	0.95	1.24
$Q_3(-)$	(Low self-sentiment)	1.02	0.80	0.87	–	0.97
Q_4	(High ergic tension)	1.23	1.04	1.11	1.14	–

Table 1 gives the corrected correlations for men (upper half) and women

---

[1] It might be argued that 'equivalence coefficients' are not the proper reliabilities to use in correction for attenuation, and that some form of split-half reliability ought to have been used. Such would not be Cattell's own view, and to give his theory the optimum opportunity to prove itself we have followed his own reasoning in our procedure. It should also be noted that there are gross differences in the empirical literature about the split-half reliabilities to be expected in relation to Cattell's 16 scales; Greif (1970) gives an interesting comparison (Table 2 in his paper) of his own findings and Cattell's. For scale A, Cattell reports a correlation of 0.82, Greif of 0.28; for scale M, the values are 0.79 and 0.21; for scale N, 0.65 and –0.04. The average reliability in Greif's study is only 0.37. Had we used these reliabilities, conclusions about the intercorrelations between scales corrected for attenuation would have been much more adverse than those actually recorded.

(lower half) between the five scales which, according to Cattell, contribute to his anxiety second-order factor. Scale H, which is also included by him, has been omitted because it also contributes to exvia–invia, the other main second-order factor, and correlates more highly with this. Out of twenty coefficients, twelve are above unity, and another three are only just below unity; this leaves five coefficients in the 0.8's (three in all) and below (two). Coefficients above unity are, of course, evidence of overcorrection, and may be due to the fact that the reliabilities were calculated on groups other than the ones furnishing the actual correlations between scales; similarly, the lower coefficients may be evidence of undercorrection. However we look at these figures, they do not justify Cattell's claim that work with the second-order factors would cause one to 'lose a lot of valuable information present initially at the primary level'. As far as one can see, practically all the information contained unreliably in the primaries is contained reliably in the second-order factor; very little information indeed is left over for contribution by the primaries.

Table 2 *Intercorrelations between Cattell's five 'exvia' scales, corrected for attenuation*

		A	E	F	Q_2 (–)	H
A	(Affecto-thymia)	–	0.44	0.66	0.91	0.69
E	(Dominance)	0.28	–	0.94	0.18	0.89
F	(Surgency)	0.61	0.85	–	1.07	1.00
Q_2 (–)	(Group adherence)	0.93	0.04	0.78	–	0.86
H	(Venturesomeness)	0.50	0.72	0.87	0.76	–

The position is not quite as clear in Table 2, which presents the intercorrelations between the five scales which according to Cattell contribute to the second-order factor exvia–invia. Only two out of twenty correlations exceed unity, and another six could be rounded up to 0.9; this leaves twelve correlations below this level. Of these, three belong to scale H (Venturesomeness) which, as already noted, also loads on the anxiety factor; as its contribution is spread over two second-order factors, its correlations for either must of course be considerably lower than unity. Of the remaining coefficients, two are very low, viz. those referring to the correlations btween E (Dominance) and Q_2 (Group adherence) in the male and female samples respectively. This would appear to be a good example of what Frenkel-Brunswik (1942) has called the principle of alternative manifestation; 'different classes of behavioral expressions were often related to one drive as alternative manifestations of that drive. . . . One drive variable may circumscribe a family of alternative manifestations unrelated to each other: the meaning of the drive concept emerges in terms of families of divergent manifestations held together dynamically or genotypically, though often not phenotypically'.

Other correlations which are well below unity are those involving A and E, and A and F. A and E are both noted by Cattell as being involved in second-order factor 3 (Pathemia); E is also involved in factor 4 (Independence), and A in factor 5 (Naturalness). F is involved in factor 8 (Superego strength). We thus find that primaries whose intercorrelations do not come up to unity when second-order factors 1 and 2 are concerned are also involved in other second-order factors; this would be impossible if all their variance were taken up by one second-order factor.

If it is permissible to draw any conclusions from these figures, it must be that they fail to support Cattell's statement quoted at the beginning of this paper; primary factors add little to the contribution made by second-order factors, with the possible exception of the 'alternative manifestations' factors contrasting extraverted attitudes leading to either leadership or group adherence. The figures given are not incompatible with a general view which would regard the primaries advocated by Cattell as random groupings of items either measuring extraversion or neuroticism, or occasionally both (i.e. the items making up his factor H). Such a view would also be compatible with the fact that several writers have found it impossible to replicate Cattell's factors in independent analyses, using both his items and his methods of analysis and rotation. The figures upon which this tentative conclusion is based are of course not very precise, for reasons already given, and the fact that several of the corrected correlations exceed unity bears witness to this. In this lack of accuracy, of course, psychology is an exact replica of physics; as Taylor *et al*. (1970) point out, 'contrary to popular opinion, physics is usually not a very exact science. . . . In some cases finding agreement to within an order of magnitude (a factor of 10) is a considerable achievement'. And the reason for this lack of exactitude is the same in both sciences: 'First, most experiments deal with a complex system in which a variety of interrelated and often poorly understood phenomena are involved. Second, the pertinent theory usually provides only an approximation based on a simplified conceptual model of the system'. In spite of the obvious inaccuracies in our calculations, the data do seem to support reasonably well the writer's conception of the relation between primary and second-order factors, and to contradict that advocated by Cattell. At the very least, the data and analyses presented seem to require some form of proof from Cattell to substantiate his contention that 'one certainly loses a lot of valuable information present initially at the primary level' in working with second-order factors only. Even restricting ourselves to two only of the eight second-order factors Cattell claims to have isolated, this just does not seem to be so.

It is interesting to speculate on the reasons for this rather strange phenomenon. Eysenck and Eysenck (1969) have drawn attention to a continuum ranging from factors which are essentially tautological (T factors) to factors which are made up of many complex and divergent items (C factors); Guilford and Eysenck, in so far as they deal with primaries, are concerned more with the former, whereas Cattell is concerned with the latter. This difference emerges also in the stress laid by these different authors on high factor loadings, leading to simple structure defined by clusters of similar items (Guilford and Eysenck), or rather on high hyperplane counts (Cattell), which are compatible with much lower factor loadings, and particularly with much lower item correlations within a given primary factor. The resulting factors are called by Cattell 'surface' (T factors) and 'source' (C factors) traits, but these terms are question-begging; there is no independent evidence to show that Cattell's factors come any closer to some truly fundamental 'source' of human behaviour, and there is some evidence that Eysenck's E and N factors do (Eysenck, 1967a). As far as our analysis goes, it seems to suggest that far from being 'source' traits, Cattell's primaries are chance aggregations of items measuring E, or N, or both, as well as possibily some other second-order factors. This conclusion fits in well with the argument presented by Eysenck and Eysenck (1969) that replicable primaries are of the 'T' type, and that replicable 'C' type factors are always of the second-order kind, at least in the non-cognitive personality field.

There is now considerable international agreement regarding the vulnerability of Cattell's system to these criticisms (Becker, 1961; Borgatta, 1962; Greif, 1970; Levonian, 1961a, b; Peterson, 1960; Timm, 1968). Greif, for instance, points out that, in his detailed analysis, out of 170 items only 28 correlate more highly with the scale they are supposed to measure than with some other scale; out of fifteen scales there are six without a single item which correlated more highly with the scale it is supposed to measure than with some other scale. He also concludes, as we have done, that 'the sixteen scales cannot by any means be regarded as functionally independent, but are relatively highly correlated'. Clearly, Cattell's hypothesis of sixteen functionally independent factors being measured by his test requires considerable support if it is to continue being accepted by test users.

References

BECKER, W. C. (1961) 'A comparison of the factor structure and other properties of the 16PF and the Guilford-Martin personality inventories', *Educ. psychol. Measur., 21*, 393–404.

BORGATTA, E. F. (1962) 'The coincidence of subtests in four personality inventories', *J. soc. Psychol., 56*, 227–44.

CATTELL, R. B., EBER, H. W. and TATSUOKA, M. G. (1970) *Handbook for the Sixteen Personality Questionnaire* (*16PF*). Champaign, Ill.: Institute for Personality and Ability Testing.

EYSENCK, H. J. (1967a). *The Biological Basis of Personality*. Springfield, Ill.: Thomas.

EYSENCK, H. J. (1967b). 'Intelligence assessment: a theoretical and experimental approach', *Br. J. educ. Psychol., 37*, 81–98.

EYSENCK, H. J. and EYSENCK, S. B. G. (1969) *Personality Structure and Measurement.* San Diego: Knapp.

FRENKEL-BRUNSWIK, E. (1942) 'Motivation and behaviour', *Genet. Psychol. Monogr., 26,* 121–265.

GREIF, S. (1970) 'Untersuchungen zur deutschen Übersetzung des 16PF Fragebogens', *Psychol. Beitr., 12,* 186–213.

HENDRICKSON, A. E. and WHITE, P. O. (1966) 'A method for the rotation of higher-order factors', *Br. J. math. statist. Psychol., 19,* 97–103.

LEVONIAN, E. (1961a). 'A statistical analysis of the 16 Personality Factor Questionnaire', *Educ. psychol. Measur., 21,* 589–96.

LEVONIAN, E. (1961b). 'Personality measurement with items selected from the 16PF Questionnaire', *Educ. psychol. Measur., 21,* 937–46.

McNEMAR, Q. (1964) 'Lost: our intelligence? Why?', *Am. Psychol., 19,* 871–82.

PETERSON, D. R. (1960) 'The age generality of personality factors derived from ratings', *Educ. psychol. Measur., 20,* 461–74.

TAYLOR, B. N., LAUGHENBERG, D. N. and PARKER, W. H. (1970) 'The fundamental physical constants', *Scient. Am., 223,* (4), 62–78.

TIMM, U. (1968) 'Reliabilität und Faktorenstruktur von Cattells 16PF Test bei einer deutschen Stichprobe', *Z. exp. angew. Psychol., 15,* 354–73.

VERNON, P. E. (1965) 'Ability factors and environmental influences', *Am. Psychol., 20,* 723–33.

9 Personality and academic attainment

N. J. Entwistle

Introduction

One of the recurrent themes to be found in educational research involves the attempt to unravel the complex determinants of academic attainment. Early work concentrated on intelligence as the explanatory variable and this is still the single most effective predictor of school achievement. Subsequently, a wide variety of research reports has drawn attention to the importance of social factors and early experiences of the child in the home. Environment affects both the development of intelligence and the level of achievement. But any attempt to understand the complete causal chain associated with school attainment must include the effect of personality on the child's work in the school.

Personality has been defined by Allport (1963) as 'the dynamic organisation within the individual of those psychophysical systems that determine his characteristic behaviour and thought'. If this definition is accepted it appears inevitable that personality will be an important determinant of classroom behaviour and ultimate scholastic success. The measurement of personality is, however, far from simple and the initial lack of success in isolating personality correlates of school attainment to a large extent can be attributed to this difficulty. The advent of pencil and paper group tests which have been carefully developed and validated has stimulated renewed interest in this research area. As a result there is now a sufficiently large literature for certain trends to emerge. Most of the British studies have used the personality inventories developed by either Eysenck or Cattell. This review thus concentrates on these factorial approaches to the measurement of personality.

Measurement of personality

Eysenck and Cattell have used different methods of identifying the factors which can be used to describe human personality. The resulting structures appear dissimilar at first sight. Papers by Burt (1965) and Adcock (1965) and a recent book (Eysenck and Eysenck, 1969a) have compared and contrasted the two approaches. Basically Eysenck had strong theoretical grounds for hypothesising two important, yet distinct, dimensions underlying human behaviour. Factor analysis designed to produce unrelated

---

Source: *British Journal of Educational Psychology*, 1972, *42*, 137–51.

factors, allowed Eysenck to choose items for his inventory which measured the traits of extraversion and neuroticism (emotional instability). Among children at least, the correlation between these two dimensions is generally found to be significantly negative and not zero. However, these scales are still being developed and new versions are expected shortly.

An understanding of the terms extraversion and neuroticism, as used by Eysenck, demands a study of the items included in the various inventories (Maudsley or Eysenck in both junior and adult versions). Neuroticism is characterized by unnecessary worrying, by feelings of restlessness, by moodiness and by general nervousness. The stable person shows behaviour which is generally controlled; he tends to be reliable, even-tempered and calm. The typical extravert is 'sociable, likes parties, has many friends, needs to have people to talk to, and does not like studying by himself. He craves excitement, takes chances, often sticks his neck out, acts on the spur of the moment, and is generally an impulsive individual. . . . The typical introvert, on the other hand, is a quiet retiring sort of person, introspective, fond of books rather than people; he is reserved and distant except with intimate friends. He tends to plan ahead, "looks before he leaps", and distrusts the impulse of the moment' (H. J. Eysenck, 1965). Scores on these two dimensions are expected to be normally distributed with the majority of people being neither extraverted nor introverted, neurotic nor stable. For certain analyses, however, it is useful to identify individuals with above or below average scores on each trait and to describe four personality 'types'—stable extraverts, stable introverts, unstable extraverts and unstable introverts.

Cattell (1965) prefers to use an alternative method of factor analysis which produces inter-related factors. Out of a vast range of items, he has isolated up to seventeen primary source traits which are given the alphabetic notation shown later in Table 1. Because the factors are inter-related it is possible to simplify Cattell's structure by repeated factor analyses which extract higher-order factors. Warburton (1968a) provided a lucid summary of the pattern of higher-order factors. The five second-order factors are reduced by successive analyses down to two fourth-order factors, as shown below:

second-order	*third-order*	*fourth-order*
stability conservatism	adaptation	integration
extraversion toughmindedness	thrust	
conscientiousness or morality	morality	morality

It is interesting to note that the two most important second-order factors are extraversion and stability, which are similar to Eysenck's major dimensions. Another two factors carry labels identical to what Eysenck sees as social attitudes (Eysenck, 1951). Eysenck has also moved away from the simple two-factor structure both by adding a psychoticism dimension (Eysenck and Eysenck, 1969b) and by subdividing extraversion into sociability and impulsivity (Eysenck and Eysenck, 1963). There is thus a considerable amount of overlap in the two descriptions of personality. In terms of the actual inventories the Cattell questionnaires (CPQ, HSPQ and 16PF for children, high-school pupils and adults) have the disadvantage of being lengthy and yet having short, rather unreliable, sub-scales. They also have American phraseology in places. Nevertheless, the possibility of being able to investigate up to seventeen dimensions, rather than two, has an obvious appeal.

Investigations into the relationship between personality and academic attainment have used both types of inventory. Taken overall the findings appear somewhat contradictory, but a pattern emerges if the studies are examined separately at each of the main educational stages.

Research at school level

Two large-scale investigations of the relationships between personality and attainment among school children have been reported. Eysenck and Cookson (1969) were able to analyse the test scores of some 4,000 pupils in their final primary year from schools throughout Staffordshire. Entwistle and Cunningham (1968) used data from an almost complete age-group of 3,000 thirteen-year-olds in Aberdeen secondary schools. Even with two such large samples there is still no close agreement between the results. An examination of the differences does, however, draw attention to some of the complexities involved in this research area.

In the introduction to their article Eysenck and Cookson provide an excellent review of the literature, including studies using both types of inventory. They conclude, from examination of studies by Jones (1960), Morrison *et al.* (1965), Ridding (1967), Rushton (1966) and Savage (1966), that the extravert tends to be better at school work in late primary and early secondary school. Similarly, the work of Butcher *et al.* (1963), Callard and Goodfellow (1962), Hallworth (1961) and Lunzer (1960) is used to hypothesize a link between neuroticism and poor performance at this age level.

In their own study Eysenck and Cookson confirmed that it is likely to be the stable extraverts, as a group, who are more successful among primary school pupils. Their zonal analyses did provide some indications on non-linearity, but the patterns of mean scores by personality type did not follow the 'Yerkes-Dodson' law: there was no tendency for both high and low scores on neuroticism to be related to poor academic results. Other studies on school children (Brown, 1970; Banks, 1964) have found this inverted-U relationship, as did Lynn and Gordon (1961) with students, but

Entwistle and Cunningham (1968) demonstrated linearity of regression among their findings.

This last study with thirteen-year-olds in Aberdeen secondary schools again found that stability was related to academic success. On the other hand, the extraversion dimensions produced an unexpected sex difference. Extraverted girls but introverted boys tended to be more successful. The difference was small, but does indicate the importance of analysing results separately by sex.

The main difference between the Aberdeen results and those of Eysenck and Cookson lies in the higher correlation between extraversion and attainment in the latter study. Eysenck and Cookson attribute this effect to the age difference in the samples and point out that there is a tendency for introversion to be associated with academic success among older children and students. This suggestion also links neatly with some of the findings based on the Cattell inventories.

The multi-dimensional approach used by Cattell makes it difficult to present a coherent picture of the results of studies using this definition of personality. The use of many more traits might, nevertheless, make the possibility of high multiple correlations more likely. Cattell, Sealy and Sweney (1966) maintain that up to 25 per cent of the variance in school attainment may be attributable to the effects of personality. This suggestion is, however, speculative as it was based on correlations corrected for the effects of unreliability in the measures. Even after these corrections the value of correlations between the primary traits and attainment remained below 0.3. Nevertheless, some agreement is to be found between this study and that reported by Cattell and Butcher (1968). The primary traits of conscientiousness (G+), submissiveness (E–), friendliness (A+) and dependability (Q_3) are related to achievement in both samples. Cattell and Butcher report a regression equation, which, by including the dimension of intelligence (B), leads to multiple correlation of 0.7 with academic success. The constants in the equation indicate the weightings given to each trait.

$$\begin{aligned}\text{Achievement} = &+ 0.15A + 0.50B + 0.10C - 0.10D - 0.15E + 0.10F\\ &+ 0.25G + 0.10H - 0.10I + 0.15J - 0.10O + 0.20Q_2 + 0.20Q_3 - 0.10Q_4\end{aligned}$$

It is too soon, however, to put much trust in such a prediction equation. The variations in relationship between samples will lead to considerable fluctuation in the weightings. In fact, individual studies taken separately do not provide a coherent picture in terms of the primary source traits. The unreliability of the sub-scales, the doubt as to the meaning underlying the individual traits (Eysenck 1971) and the small scale of many of the studies, all combine to make this confusion inevitable. Fortunately, a lucid summary of research using the Cattell tests is already in existence (Warburton, 1968b) and has been particularly valuable in preparing this paper.

Warburton, who was closely associated with the work of Cattell for many years, chose twenty-nine of the most satisfactory studies to demonstrate the existence of a pattern of relationships in terms of the second-order factors. Table 1 presents a condensation of Warburton's summaries, based on the twenty-five studies which used Cattell inventories. The articles from which these results were derived are shown among the references, marked with asterisks.

		Age Group							
Second-order Factor	*Primary Source Trait*	*Under 12*		*12–14*		*15–17*		*Over 17*	
		+	–	+	–	+	–	+	–
Anxiety/ Neuroticism	Instability (C–)	0	3	0	1	3	0	3	2
	Tenseness (Q_4+)	0	1	0	1	1	0	5	1
	Insecurity (O+)	0	0	0	2	1	0	1	1
	Bad-temperedness (D+)	0	1	0	2	2	0	0	0
	Unreliability (Q_3–)	0	1	0	5	0	2	0	2
	Suspiciousness (L+)	0	0	0	0	0	0	2	2
	Total	0	6	0	11	7	2	11	8
Introversion (1)	Aloofness (A–)	0	0	0	6	3	0	0	1
	Desurgency (F–)	0	3	0	0	8	0	3	1
	Shyness (H–)	0	2	0	2	3	0	1	0
	Total	0	5	0	8	14	0	4	2
Introversion (2)	Submissiveness (E–)	0	1	7	0	4	0	1	2
	Independence (Q_2+)	1	0	3	0	3	0	1	0
	Total	1	1	10	0	7	0	2	2
Tendermindedness	Tendermindedness (I+)	0	0	1	0	2	0	3	0
	Unaffectedness (N–)	0	0	0	0	0	0	1	0
	Total	0	0	1	0	2	0	4	0
Radicalism	Bohemianism (M+)	0	0	0	0	0	0	4	0
	Radicalism (Q_1+)	0	0	0	0	0	0	1	0
	Total	0	0	0	0	0	0	5	0
Morality	Conscientiousness (G+)	2	0	6	0	4	1	2	3

The signs following the letters in parentheses indicate the direction of scoring on the trait. For example, shyness is inferred from low scores (negative sign) on trait 'H'.

Table 1 *Summary of results from studies using Cattell's inventories: direction of correlation coefficients with scholastic attainment (adapted from Warburton, 1968b, table 3).*

Warburton drew attention to the same age trend noted by Eysenck. Stable extraversion is related to success in the primary school, changing over to introversion and possibly to emotional instability by age eighteen. From the pattern shown in Table 1 it appears that the change-over both from stability to neuroticism and from extraversion to introversion in fact occurs late in the secondary school. The abrupt reversal of relationships after age fourteen, particularly that between introversion and attainment, is quite remarkable. All thirteen relationships obtained with younger children are negative, with extraverts being more successful; out of eighteen with older pupils and students only two are in this direction. Perhaps the change-over from junior to senior versions of the inventories affects the results, but it is difficult to explain such a clear-cut change in the direction of the relationship in this way. The fact that this reversal coincides with the school leaving age may give an indication as to its cause. Between fifteen and eighteen, the majority of young people move into employment and out of the reach of educational research workers. The remaining captives are highly selected on both intelligence and social class. It may be that the different relationships between personality and attainment are a characteristic of this restricted population found in sixth forms and in colleges and universities.

Research at college and university level

Warburton's summary provides convincing evidence of the superiority of the introvert in higher education. Savage (1962) in Australia and Lavin (1967), who reviewed the American literature, both reached the same conclusion. The precise relationship of stability-instability to good performance is, however, much less clear. Furneaux (1962) suggested the importance of instability; Kelvin, Lucas and Ojha (1965) showed that the first-class honours graduates and those students who failed both showed high neuroticism scores on the whole. But other studies by Savage (1962) and by Lynn and Gordon (1961) have found a different relationship in which students who had high or low instability scores did worse than those who had moderate scores. The American literature also produces contradictory evidence, in spite of the use of large samples in some studies. Lavin (1967) concludes that, on balance, the stable student is likely to be the more successful. Warburton and Eysenck both reached the opposite conclusion in Britain, but results from recent studies have been more equivocal. Kline and Gale (1971) found little evidence of a significant relationship between either neuroticism or introversion and marks in a first-year psychology examination. Wankowski (1968) has suggested that it is stable introverts who do best at university. Smithers and Batcock (1970) find similar results for social scientists in a technological university, but not for health scientists. Variations in relationships among students taking different academic disciplines, are also reported in a recent study by Wilson (1971).

Extraversion was significantly related to failure for all students in the arts faculty. Stability was linked with failure, but only for arts women. This study also shows that factors indicating success are not necessarily the converse of those related to failure. Equations predicting success may include rather different variables than those used in identifying students who are at risk.

Not only are different relationships to be expected in contrasting subject areas, but there may also be differences among samples drawn from the three distinct sectors of higher education. Most of the evidence comes from universities. To date no consistent pattern of results seems to be emerging from studies either in technical colleges (Staveley, 1967; Venables, 1967; Cowell and Entwistle, 1971) or in colleges of education (Halliwell, 1963; Cortis, 1966). Clearly it is important to have results in which the same tests have been given to comparable samples drawn from all three sectors. A study at Lancaster is following-up a group of some 2,600 students from twenty different institutions. A preliminary analysis on about half of this sample (Entwistle *et al*., 1971b) has already shown interesting differences both between areas of study and, to a lesser extent, between universities and colleges. For example, neurotic introverts apparently make good engineers, as Furneaux found, but stable introverts do better in the pure sciences. Neurotic introverts are also good at languages, but stable introverts make better historians. Among science students the correlations between both personality dimensions and attainment were near zero, which fits in with the findings of Kline and Gale.

Taken overall the greatest consistency in results occurs with the extraversion trait. On the whole introverts do tend to be more successful. Correlations ranged from –0.26 to –0.03 in different subject areas at university, from –0.21 to –0.07 in comparable groups at college of education and from –0.10 to 0.08 at polytechnics. The neuroticism trait was much less consistent, equivalent ranges being –0.24 to 0.28, –0.27 to –0.07 and –0.42 to –0.05. Neuroticism is positively correlated with academic performance only in certain sub-groups of the student population, at least on the evidence of this particular study.

There thus appears to be some doubt about the generality of the earlier suggestion that unstable students tend to be successful. The findings are conflicting and, while the possibility of subject-area differences may help to resolve some of the confusion, there are likely to be alternative explanations for this situation. For example, Eysenck (1972) suggests that it is necessary to distinguish between trait anxiety, which is usually measured, and state anxiety, which may be more relevant in determining examination success. Eysenck goes on to attack the use of the simple Hullian equation relating performance to drive and habit. It is important to identify habit-systems which are important in academic behaviour. Under this heading might be included study methods, attitudes to studying, and the dimension recently described by Hudson (1968) as 'syllabus-bound/syllabus-free'. Academic motivation is perhaps in an intermediate position between drive and habit. Eysenck suggests that 'theoretically at least, introverted habit systems would seem likely to pre-dispose students to engage in good study habits, so that high anxiety drive in introverts would lead to even more strenuous study'. It is not known at present whether good study habits combined with introversion and neuroticism are particularly likely to lead to academic success, but empirical results do indicate that good study habits, as defined in two separate inventories, are related to stable introversion (Entwistle and Wilson, 1970; Entwistle and Entwistle, 1970; Cowell and Entwistle, 1971; Entwistle *et al.*, 1971a). Also academic motivation is linked to stability (Entwistle, *et al.*, 1971b) and the syllabus-bound dimension to introversion (John, 1972).

The link between stability (*low* drive) and either motivation or study methods is not ruled out in Eysenck's elaboration of the elementary theory of drive and performance. However, it is possible to put forward alternative formulations of the effect of neuroticism on academic study. It may be precisely the combination of high motivation with emotional stability which summarizes the characteristics of the student with 'high drive'—at least in the everyday use of this term. Descriptions of stability imply low reactivity to stimuli. Perhaps in this particular context we need to extend the neuroticism dimension further to include strong reaction in a non-neurotic direction. Then drive might be linked with high reactivity in either direction and both extremes might lead to success for differing groups of students. This would represent a parallel to the distinction met elsewhere in the literature between 'achievement motivation' and 'fear of failure' (Birney, Burdick and Teevan, 1969). The Eysenckian formulation, which modifies the direction of anxiety (drive) according to previous habits, may well coincide with the concepts of 'facilitating' and 'debilitating' anxiety as described by Alpert and Haber (1960). It certainly appears that the use of elaborate theoretical models may lead to more sophisticated investigations into the relationship between personality and attainment; there is also room for improvement in the research designs used in such studies.

Problems and growth-points

There is a variety of factors which may explain the contradictions between findings which abound in the literature reviewed in this paper. Before attempting a final summary of the outcomes of this research, various aspects of research design which may have affected results will be briefly discussed.

Methods of analysis

The commonest forms of analysis applied to data on personality and attainment have been correlation techniques and zonal analysis. Both approaches lead to difficulties. The major problem involved in zonal analysis is the definition of the various personality 'types'. Many studies have created four types simply by dichotomizing the scores on each dimension at the median value. To equalize the size of groups it is better to plot the two-dimensional scatter diagram of extraversion against neuroticism. But both these methods are specific to a particular sample and prevent direct comparisons being made between studies. Agreement to define personality types in terms of the norms presented in the test manuals for each age-level or sub-group of the adult population would be a useful move, at least where the norms have been derived from samples of adequate size. As much of the research is done on rather small groups, it might be necessary to make the extreme groups fairly close to the overall mean. Personality types might then be created by taking scores of $\pm 0.5\ \sigma$ either side of the means reported in the manuals, or, where necessary, dichotomizing at this mean value. Studies having larger samples might be able to report additional analyses of more extreme groups (perhaps $\pm 0.75\ \sigma$). Nevertheless, the advantage of comparability produced by a standard way of creating 'types' is off-set by the artificiality and rigidity of using arbitrary cut-off points which do not reflect, even in an intuitive way, variations in relationships existing in the data. Clearly other approaches and other methods of analysis should also be attempted.

Most of the research reported in this paper has involved the use of various correlational techniques. While these methods of analysis have distinct advantages in terms of ease of communication, they do have definite weaknesses. It is possible to dismiss a non-linear relationship in reporting an insignificant product-moment coefficient and also to ignore the possibility of important interactions between variables. Correlations average out the relationships over the whole sample, ignoring possible differences between sub-groups. It is certainly useful to apply correlational techniques to sub-groups, where these are known in advance. But, in view of possibilities such as the one discussed in the previous section, that anxiety may affect the performance of different students in opposite directions, alternative methods of analysis might be more appropriate.

Whereas correlational techniques compare similarities between *tests* over the whole sample, the method of cluster analysis and classification procedures compare similarities between *people* over the complete battery of tests. They provide indications of groups of people who have the same patterns of scores. It would thus be possible, in principle, to identify two separate groups of students both with high neuroticism scores, one showing high and the other low attainment scores. Recent developments in computers have made cluster analysis a more attractive possibility and some promising analyses of data have already been reported (Entwistle and Brennan, 1971). One classification procedure in particular (automatic interaction detection) allows sub-groups to be produced on the basis of the various interactions existing in the data, rather than being created by arbitrary or subjective decisions. This method of analysis seeks to maximize the prediction of a criterion variable by forming sub-groups in terms of a small number of 'explanatory' variables. It thus appears to be a potentially fruitful approach to the problem of isolating interactions between personality and other variables in relation to academic performance. Early results from an analysis of data from university students (Brennan and Entwistle, 1972) have thrown up the possibility that introversion is related more closely to attainment in a sub-group of science students with moderate A-level results who see themselves as being 'hard-working', than in other sub-groups.

Measurement of attainment

The results reported in earlier sections were all based on scores from either Eysenck or Cattell personality inventories, but academic attainment was measured in a variety of ways. Some of the differences in findings may be attributable to these variations.

In some studies attainment is measured by objective tests of English and mathematics; in others verbal reasoning tests are used. A few studies have reported analyses based on performance in programmed learning tasks, in which the time allowed for these tasks may be an important consideration in itself (Seddon, 1972). External or internal examination marks are found frequently in the literature. While all these measures may well show fairly high intercorrelations, they still represent alternative definitions of attainment, which could have rather different personality correlates. 'Academic attainment' surely implies that pupils or students have been assessed in terms of their performance on the courses they have actually taken in school or college. Thus, tests of verbal reasoning, though correlated with academic attainment, are not direct measures of success in school. Similarly objective tests of English and mathematics provide only two components of school attainment. Even in primary schools many other mental skills are important–for example, there is an increasing emphasis on individual expression in music, art and language. It seems improbable that a single set of personality correlates will adequately describe success in such different forms of intellectual activity.

There is some evidence that different personality characteristics are linked to success with contrasting definitions of attainment (W. Rudd, 1970; Bennett and Entwistle, 1972). Bennett, in a piece of ongoing research, is using two contrasting measures of attainment in primary, middle and secondary schools. Early results from the primary stage are based on a conventional measure of school attainment (teachers' rank-orders scaled on a verbal reasoning test) and on impression marking of imaginative stories. As anticipated, rather different relationships between personality and attainment are emerging from the two types of measure. Bennett has confirmed the superiority of stable extraverts in conventional attainment at this age-level, but a sex difference was found when results from the imaginative stories were analysed. High scores were given to boys who were stable extraverts, but to girls who were unstable extraverts.

Similar differences might also be expected in secondary schools. Again standardized tests, used on their own, may not correspond at all closely to the teacher's view of school attainment. With a large number of subjects included in the curriculum perhaps the most valid measure of overall achievement will be a sum or an average of marks taken from the whole range of subjects. These marks will generally have to be scaled against a standardized test of scholastic aptitude to allow comparisons between classes and between schools. The assumption here is that the consensus of opinion among teachers provides the best indication of the rank-order of achievement within a class, while an objective test is a more accurate way of deciding the standard of that class compared with others. During the later stages of secondary school sums (or averages) of grades at O- or A-level in the GCE make satisfactory measures of academic success without any scaling procedure.

Finally, in higher education the growing tendency to include both course work and examination marks in the overall assessment of students suggests further division in the definition of attainment. Differences in the correlates of success from discipline to discipline have already been demonstrated. There is a distinct possibility that course work and examination may favour different personality types. Again research on this possibility is already being undertaken (E. Rudd, 1970).

Additional complications

Throughout this review it has been assumed that the measurements of extraversion and neuroticism using Eysenck's scales are internally consistent and that the scores are stable. The manual (S. B. G. Eysenck, 1965) reports split-half reliabilities varying from age to age around 0.8 and test/re-test reliabilities after one month around 0.7. Only recently have follow-up studies over a longer period given further evidence on the degree of stability to be expected. Sharples (1971) with primary school children has found correlations on the JEPI over one year to be down to about 0.55. Bennett (1972) has summarized evidence from a number of sources. One analysis produced test/re-test coefficients of 0.38(E) and 0.51(N) after one year (ages 10–11), while other results compared scores on the JEPI and the EPI. When the tests were taken on the same occasion by fifteen-year-olds, correlations of 0.82(N) and 0.87(E) were obtained. In spite of this high level of overlap between the senior and junior versions, the correlation between equivalent scales from age thirteen to fifteen was found to be down to 0.33. If these values are substantiated, they would suggest a degree of instability previously unsuspected in the literature. Shifts in relationships between personality and attainment with age might then be explained as much in terms of changes in personality within contrasting academic environments as in changes in attainment.

Bennett (1972) also pointed out some apparent weaknesses in the extraversion scale of the JEPI. Response rates of over 80 per cent are usually viewed with concern in the development of Likert-type tests. In three separate age groups Bennett found the same twelve extraversion items with response rates above this level, and these items produce a lack of 'ceiling' on this scale. Looking at the items it appears that the 'translation' of adult situations from the EPI into the JEPI may have led to anomalies. Most of the twelve items refer to social extraversion and three refer specifically to behaviour at parties. Such social occasions presumably have a different meaning for adults than for children; a translation of situations does not necessarily make the items equally useful for different populations. It seems possible that impulsivity items, rather than social extraversion items, may be distinguishing between extraverts and introverts in the JEPI. This situation might again affect the observed relationships between personality and attainment.

Another complication relates to the possibility of various intervening variables which may affect the size or the direction of the relationship between the two main dimensions under consideration. It is already clear that sex and age have been acting in this way. Further analysis of the data obtained from the follow-up of Aberdeen secondary pupils by Entwistle and Welsh (1969) suggested that there might also be different relationships among bright children compared with those found for pupils of below average ability. Dividing their sample into top and bottom thirds in terms of ability level, they found that introversion was linked to high achievement for bright boys, whereas extraversion was the more favourable trait in the low ability group. The results for girls were similar, but less marked.

Some support for this finding can be obtained by taking two comparable studies with secondary modern children on the one hand (Fraser, 1967) and with grammar school pupils (Duckworth and Entwistle, 1972) on the other. Both studies used the JEPI with children aged 12–13 years and obtained separate correlations with attainment in English and Mathematics. Fraser's secondary modern sample showed positive correlations between extraversion and both measures of school attainment. The grammar school sample produced correlations which were both negative. It does seem possible, although the evidence is still tenuous, that introversion is related to success for the brightest children in secondary school even as early as the second year. Among the low ability group extraversion is still associated with high achievement at age fourteen (Entwistle and Welsh, 1969).

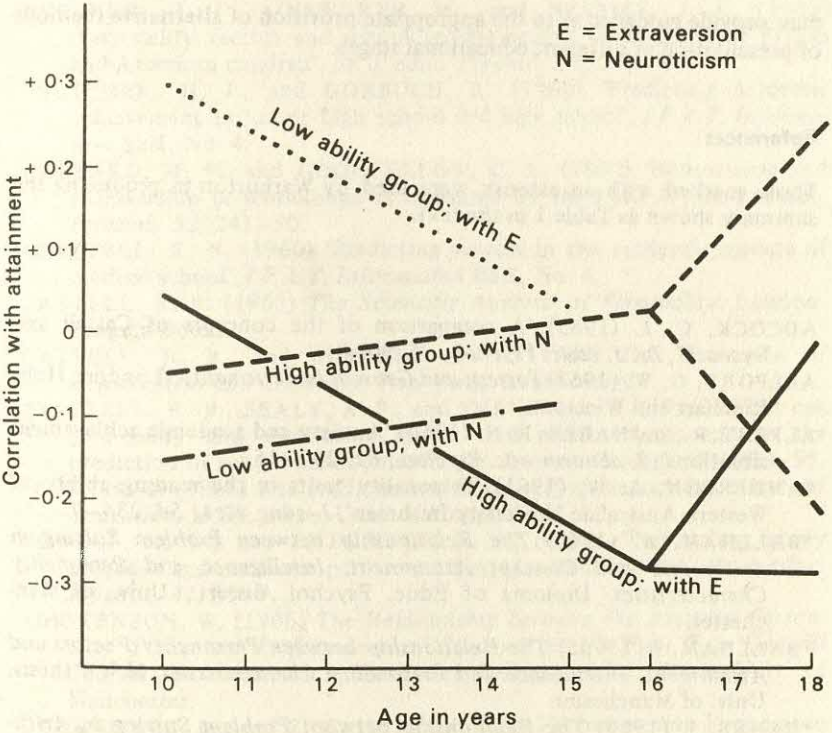
There are other intervening variables which may affect the relationships between personality and attainment. A number of studies have observed differing relationships among urban children and rural children (Cattell and Butcher, 1968; Entwistle, 1967), but these may be a reflection of differences in classroom organization, in average class size or in methods of teaching. The range of approaches in primary schools, between formal and informal methods, suggests that different personality types would be successful in contrasting educational climates. These possibilites are being investigated at present (Sharples, 1971; Bennett and Entwistle, 1972; Eysenck, 1971).

Finally, it seems probable that there will be an interaction between the teacher's personality and the performance of children of differing personality types. There are formidable obstacles to research of this kind, but it may be crucial to overcome them. It may be only in this way that the observed relationships between personality and attainment will be fully explained.

Conclusion

The last few sections have outlined a personal view of problems and growth-points in research into personality and attainment. Recent work at Lancaster has figured rather prominently in this subjective perspective, as the author is less familiar with ongoing research elsewhere. There are almost certainly important developments which have been undervalued or ignored altogether. Similarly the map drawn in describing the literature may represent the operation of selective perception. Despite this, two major conclusions may be drawn with some confidence. First, it is dangerous to assume wide generality in statements about the relationships between personality and academic attainment. Age, ability, sex, geographical area, classroom organization, class size, teaching methods and teacher personality may all affect these relationships to some extent. Nevertheless, the second conclusion does indicate the existence of some overall pattern in this confused research area: there is now considerable agreement as to the existence of an age effect. Many apparent contradictions in the literature are resolved by the recognition that stable extraverts tend to be successful in primary schools, while introverts, and possibly even neurotic introverts predominate among outstanding students. It is not clear, however, to what extent personality, rather than attainment, is changing over this period.

Figure 1 represents an intuitive graph, based partly on research evidence, which indicates how the pattern of observed relationships may change with the age and ability level of the pupils forming the sample. Extraversion and stability remain advantageous longer for children of below-average ability. Introversion is related to success among bright children soon after entering secondary school. It may be possible to simplify even further the relationships shown in Figure 1. If the correlation were to be plotted against *mental age* rather than chronological age, a single line might adequately describe each of the relationships. If a bright child of age ten can be assumed to be equivalent intellectually to a child of low ability aged about fifteen, the four discrete lines would almost merge into two. However, such



Variations after age sixteen dependent on area of specialisation.

Fig. 1 Indication of variations in correlations between personality (E and N) and conventional measures of academic attainment, by age and ability.

neatness is unlikely to coincide with reality in this particular research area.

Finally, it may be salutary to look for practical, rather than theoretical, importance in the research findings. It still requires considerable faith and imagination to see where these results may lead. The lack of generality in relationships and the probable link between teaching methods and personality correlates of success both reinforce the growing realization of the importance of variety in the presentation of material to be learned. Even at the same intellectual level no one teaching method or way of organizing learning experiences will be ideal for all pupils or for all students. Ultimately, it may be possible to allow the 'consumers' to choose the method of instruction, at least in the later stages of education. Formal, systematic presentation may still be preferred by some; unstructured freedom may suit others. An understanding of personality differences may help to anticipate such preferences. Research into the relationships between personality and academic attainment, with an emphasis on changes with age, may provide guidance as to the appropriate provision of alternative methods of presentation at different educational stages.

References

Those marked with an asterisk were used by Warburton in producing the summary shown as Table 1 in the text.

ADCOCK, C. J. (1965) 'A comparison of the concepts of Cattell and Eysenck', *Br. J. educ. Psychol., 35,* 90–8.

ALLPORT, G. W. (1963) *Pattern and Growth in Personality.* London: Holt, Rinehart and Winston.

ALPERT, R., and HABER, R. N. (1960) 'Anxiety and academic achievement situations', *J. abnorm. soc. Psychol., 61,* 207–15.

*ANDERSON, A. W. (1961) 'Personality traits in the reading ability of Western Australian University freshman', *J. educ. Res., 54,* 234–7.

*BALLHAM, A. (1960) *The Relationship between Problem Solving in Arithmetic and Concept Attainment, Intelligence and Personality Characteristics.* Diploma of Educ. Psychol. dissert., Univ. of Manchester.

*BALLHAM, A. (1965) *The Relationship between Personality Factors and Attainment, Intelligence and Personality Characteristics.* M.Ed. thesis, Univ. of Manchester.

*BANKS, J. (1964) *The Relationship between Problem Solving in Arithmetic and Concept Attainment, Intelligence and Personality Characteristics.* M.Ed. thesis, Univ. of Manchester.

*BEECH, F. (1963) *Investigation into Effects of Personality Factors on Progress in English and Mathematics in Two Secondary Modern Schools.* Unpublished D.A.S.E. dissert., Univ. of Manchester.

BENNETT, S. N., and ENTWISTLE, N. J. (1972) *Correlates of Creative and Conventional Attainment in Primary School.* (In preparation.)

BENNETT, S. N. (1972) *Similarity of Response Pattern on the J.E.P.I. in relation to Conventional Creative Attainment.* Paper read at the B.P.S. Annual Conference at Nottingham.

BIRNEY, R. C., BURDICK, H., and TEEVAN, R. C. (1969) *Fear of Failure.* New York: Van Nostrand.

BRENNAN, T., and ENTWISTLE, N. J. (1972) *Predicting Academic Performance of Arts and Science Students in Multivariate Classification.* (In preparation.)

BROWN, G. (1970) 'An investigation into the relationships between performance and neuroticism', *Durham Res. Rev., 25,* 483–8.

*BURDICK, L. A. (1963) 'Analysis of sixteen personality factor questionnaire and elementary student teachers at Indiana State College', *College Journal,* 1963, 57–9.

BURT, C. (1965) 'Factorial studies of personality and their bearing on the work of the teacher', *Br. J. educ. Psychol., 35,* 368–77.

BUTCHER, H. J., AINSWORTH, M., and NESBITT, J. E. (1963) 'Personality factors and school achievement–a comparison of British and American children', *Br. J. educ. Psychol., 33,* 276–85.

*BUTCHER, H. J., and GORSUCH, R. (1960) 'Predicting academic achievement in junior high school and high school', *I.P.A.T. Information Bull.,* No. 4.

CALLARD, M. P., and GOODFELLOW, C. L. (1962) 'Neuroticism and extraversion in schoolboys as measured by the J.M.P.I.', *Br. J. educ., Psychol., 32,* 241–50.

*CATTELL, R. B. (1960) 'Predicting success in the academic aspects of medical school', *I.P.A.T. Information Bull.,* No. 4.

CATTELL, R. B. (1965) *The Scientific Analysis of Personality.* London: Penguin Books.

CATTELL, R. B., and BUTCHER, H. J. (1968) *The Prediction of Achievement and Creativity.* New York: Bobbs-Merrill.

*CATTELL, R. B., SEALY, A. P., and SWENEY, A. B. (1966) 'What can personality and motivation source trait measurements add to the prediction of school achievement?', *Br. J. educ. Psychol., 36,* 280–95.

*CORTIS, G. (1966) *The Prediction of Student Performance at Colleges of Education.* M.Ed. thesis, Univ. of Manchester.

COWELL, M. D., and ENTWISTLE, N. J., (1971) 'Personality, study attitudes and academic performance in a technical college', *Br. J. educ. Psychol., 41,* 85–9.

*DICKENSON, W. (1966) *The Relationship between the Abilities, Personality Traits and Performance of Students in the First Year General Courses for Engineers in Technical Colleges.* M.Ed. thesis, Univ. of Manchester.

DUCKWORTH, D., and ENTWISTLE, N. J. (1972) *Intervening Variables in the Relationship between Personality and School Attainment.* Unpublished report, Department of Educational Research, University of Lancaster.

ENTWISTLE, N. J. (1967) *The Transition to Secondary Education: The Age of Transfer and Correlates of Academic Success in the First Year of Secondary School.* Unpublished Ph.D. thesis, Univ. of Aberdeen.

ENTWISTLE, N. J., and BRENNAN, T. (1971) 'The academic performance of students, II–Types of successful students', *Br. J. educ. Psychol., 41,* 268–76.

ENTWISTLE, N. J., and CUNNINGHAM, S. (1968) 'Neuroticism and school attainment–a linear relationship?', *Br. J. educ. Psychol., 38,* 123–32.

ENTWISTLE, N. J., and ENTWISTLE, D. M. (1970) 'The relationships between personality, study methods and academic performance', *Br. J. educ. Psychol., 40,* 132–43.

ENTWISTLE, N. J., NISBET, J. B., ENTWISTLE, D. M., and COWELL, M. D. (1971a) 'The academic performance of students, I–Prediction from scales of motivation and study methods', *Br. J. educ. Psychol., 41,* 258–67.

ENTWISTLE, N. J., PERCY, K. A., and NISBET, J. B. (1971b) *Educational Objectives and Academic Performance in Higher Education.* Unpublished report, Dept. of Educational Research, Univ. of Lancaster.

ENTWISTLE, N. J., and WELSH, J. (1969) 'Correlates of school attainment at different ability levels', *Br. J. educ. Psychol., 39,* 57–63.

ENTWISTLE, N. J., and WILSON, J. D. (1970) 'Personality, study methods and academic performance', *Univ. Quart., 24,* 147–56.

EYSENCK, H. J. (1951) 'Social attitudes questionnaire and scoring key', *Br. J. Psychol., 42,* 114–22.

EYSENCK, H. J. (1965) *Fact and Fiction in Psychology.* London: Penguin Books.

EYSENCK, H. J. (1971) 'On the choice of personality tests for research and prediction', *J. Behav. Sci. 1,* 85–9.

EYSENCK, H. J. (1972) 'Personality and attainment: an application of psychological principles to educational objectives', *Higher Educ., 1,* 39–52.

EYSENCK, H. J., and COOKSON, D. (1969) 'Personality in primary school children: I–Ability and achievement', *Br. J. educ. Psychol., 39,* 109–22.

EYSENCK, H. J., and EYSENCK, S. B. G. (1969a) *Personality Structure and Measurement.* London: Routledge and Kegan Paul.

EYSENCK, H. J., and EYSENCK, S. B. G. (1969b) ' "Psychoticism" in children: a new personality variable', *Res. in Educ., 1,* 21–37.

EYSENCK, S. B. G. (1965) *J.E.P.I. Manual.* London: University of London Press.

EYSENCK, S. B. G., and EYSENCK, H. J. (1963) 'On the dual nature of extraversion', *Br. J. soc. clin. Psychol., 2,* 46–55.

FRASER, B. (1967) *An Investigation into the Relationship between Personality Variables and Attainment in a Secondary Modern School.* Unpublished dissert. (Diploma in Educ. Psychology–Child Guidance), Univ. of Manchester.

FURNEAUX, W. A. (1962) 'The psychologist and the university', *Univ. Quart., 17,* 33–47.

*HALLIWELL, C. (1963) *The Relationship between Personality Measures and Performance in a Training College for Teachers.* Ph.D. thesis, Univ. of Sheffield.

HALLWORTH, H. J. (1961) 'Anxiety in secondary school children', *Br. J. educ. Psychol., 31,* 281–91.

*HOLLAND, J. L. (1960). 'The prediction of college grades from personality and aptitude variables', *J. educ. Psychol., 50,* 245–54.

*HOLLAND, F. J. (1960) 'Predicting academic success on a general college curriculum', *I.P.A.T. Information Bull.,* No. 4.

HUDSON, L. (1968) *Frames of Mind.* London: Methuen.

JOHN, K. (1972) *Some Differences in the Personality, Attitudes and Cognitive Styles of Students Choosing Arts or Science Specialisms in the Sixth Form and at University.* Unpublished M.Sc. thesis, Univ. of Lancaster.

JONES, H., GWYNNE (1960) 'Relationship between personality and scholastic attainment', *Bull. Br. Psychol. soc., 40,* 42.

KELVIN, R. P., LUCAS, C., and OJHA, A. (1965) 'The relation between personality, mental health and academic performance of university students', *Br. J. soc. clin. Psychol., 4,* 244–53.

*KLINE, P. (1966) 'Extraversion, neuroticism and academic performance among Ghanaian university students', *Br. J. educ. Psychol., 36,* 93–4.

KLINE, P., and GALE, A. (1971) 'Extraversion, neuroticism and performance in a psychology examination', *Br. J. educ. Psychol., 41,* 90–3.

LAVIN, D. E. (1967) *The Prediction of Academic Performance.* New York: Wiley, Science Editions.

*LOCKE, M. (1958) *The Relationship between Academic Performance of University Students and Various Cognitive and Personality Variables.* Unpublished report.

LUNZER, E. A. (1960) 'Aggressive and withdrawing children in the normal school', *Br. J. educ. Psychol., 30,* 119–23.

LYNN, R., and GORDON, I.E. (1961) 'The relationship of neuroticism and extraversion to intelligence and educational attainment', *Brit. J. educ. Psychol., 31,* 194–203.

*MAYFIELD, A. (1966) *Personality, Sportsmanship and Sociometric Status in Relation to Physical Ability in Technical High School.* Unpublished report.

MORRISON, A., MACINTYRE, D., and SUTHERLAND, J. (1965) 'Teachers' personality ratings of pupils in Scottish primary schools', *Br. J. educ. Psychol., 35,* 306–19.

*PORTER, R. B. (1964) 'A comparative investigation of the personality of sixth grade gifted children and a norm group of children', *J. educ. Res.,* 132–4.

RIDDING, L. W. (1967) 'An investigation of the personality measures associated with over- and under-achievement in English and Arithmetic', *Br. J. educ. Psychol., 37,* 397–8.

RUDD, E. (1970) 'Student performance in university assessments in relation to individual characteristics', *Social Science Research Council, Annual Report,* 1969–70. London: HMSO.

RUDD, W. G. A. (1970) *C.S.E.: A Group Study Approach to Research and Development.* London: Evans/Methuen Educational.

*RUSHTON, J. R. (1966) 'The relationship between personality characteristics and scholastic success in 11-year-old children', *Br. J. educ. Psychol., 36,* 178–84.

*RUSHTON, J. R. (1968) *A Longitudinal Study of Personality Characteristics and Academic Achievement.* Ph.D. thesis, Univ. of Manchester.

SAVAGE, R. D. (1962) 'Personality factors and academic performance', *Br. J. educ. Psychol., 32,* 251–3.

SAVAGE, R. D. (1966) 'Personality factors in academic attainment in junior school children', *Br. J. educ. Psychol., 36,* 91–4.

SEDDON, G. M. (1972) *Eysenckian Theory and its Relationship to Performance on Complex Tasks.* (Personal communication.)

SHARPLES, D. (1971) *A Longitudinal Study of the Personality and the Attainments and Attitudes of Junior School Children.* Unpublished Ph.D. thesis, Univ. of Bath.

SMITHERS, A. G., and BATCOCK, A. (1970) 'Success and failure among social scientists and health scientists at a technological university', *Br. J. educ. Psychol., 40,* 144–53.

*SOLOMONS, E. (1961) *An Investigation into the Relationships between*

Personality Traits, Interests and Attainments with Particular Reference to the Difficulty of the task. Diploma in Educ. Guidance Dissert., Univ. of Manchester.

*SOLOMONS, E. (1965) *Personality Factors and Attitudes of Mature Training College Students.* M.Ed. thesis, Univ. of Manchester.

*STAVELEY, B. (1967) *The Abilities and Interests of Craft and Technical Students of Mechanical Engineering.* M.Ed. thesis, Univ. of Manchester.

*VENABLES, E. (1963) 'Social differences among day release students in relation to their recruitment and examination success', *Br. J. soc. clin. Psychol., 3,* 138–52.

VENABLES, E. (1967) *The Young Worker at College–A Study at a Local Technical College.* London: Faber and Faber.

WANKOWSKI, J. A. (1968) 'Some aspects of motivation in success and failure at university.' In *Proc. Fourth Annual Conference for Society for Research into Higher Education.* London: S.R.H.E.

WARBURTON, F. W. (1968a) 'The assessment of personality traits.' In MORRIS, J. F., and LUNZER, E.A. *Development in Learning, Volume 3–Contexts of Education.* London: Staples.

WARBURTON, F. W. (1968b) *The Relationship between Personality Factors and Scholastic Attainment.* Unpublished report, Department of Education, Univ. of Manchester.

*WARBURTON, F. W., BUTCHER, H. J., and FORREST, G. M. (1963) 'Predicting student performance in a university department of education', *Br. J. educ. Psychol., 33,* 68–79.

WILSON, J. D. (1971) 'Predicting levels of first year university performance', *Br. J. educ. Psychol., 41,* 163–70.

Section Three
Motivation and learning

Introduction

The articles in this section fall into two main groups: the first two are concerned with theories of motivation and learning and with the evidence that led to their formulation, while the second group of articles describes attempts to put theory into practice within the classroom.

The first article, by Hunt, gives an overview of some of the dominant theories of motivation and learning, including Hull's drive-reduction theory and Skinner's theory of reinforcement. He then goes on to point out the weaknesses of these attempts to explain human behaviour: that organisms do behave in the absence of drive, and that behaviour does occur even when no external reward is offered. Hunt puts forward an alternative theory of motivation and learning based on Hebb's principle of the optimum level of stimulation. In Hunt's model, information-processing becomes the basic premise on which the theory is built. Organisms are motivated to learn in order to resolve the incongruity between new stimuli and previous experience.

Bruner, in his article, extends this idea and discusses how organisms, because of the incongruity between their previous experience and incoming stimuli, are intrinsically motivated to learn.

The second group of articles looks at the problem of how, using the theories discussed, we can increase a child's motivation to learn. Two different environments play their part in this process: the home and the classroom. In the first article Swift hypothesizes that there are two 'types' of home environment which are likely to produce academically successful children; the classic middle-class background which values education for its liberating qualities for the individual, and the upper-working-class background which views education as socially liberating.

The remaining three articles describe how the teacher can manipulate the environment within the classroom to increase each individual child's capacity to learn. Skinner discusses how schedules of reinforcement can be used in teaching, and the article by Schaeffer, Harris and Greenbaum provides a case study in the use of these techniques with a socially oriented underachiever. Finally, Karnes, Teska and Hodgins discuss four different preschool programmes and show that the one which structured the environment to provide the optimum amount of incongruity between the child's previous experience and the new incoming stimuli was the most successful in increasing the child's intellectual and verbal ability.

Joan M. Whitehead

10 Experience and the development of motivation: some reinterpretations

J. McV. Hunt

A recent issue of the *Saturday Evening Post* carried a cartoon that some of you may have noted. It depicts a boy entering his house, perhaps from school, where his father is sitting with his paper. The boy appears to be fixing his father with an accusing glare. The punch line reads, 'Somebody goofed, I'm improperly motivated'.

This cartoon depicts the vantage point from which I have been examining what we think we know about the relation between experience and motivation. When a child's behaviour fails to fit the standards somebody in our society holds for him, it is pretty well agreed among us who are supposed to be experts on human nature that 'somebody goofed'. And that somebody is usually considered to be a parent.

The question is: what is the proper formula? If one examines the accruing evidence relevant to what has been the dominant conception of the experiential sources of motivation, one can hardly escape the conclusion that this conceptual scheme needs some revisions. If we based our child-rearing entirely on our dominant theory of motivational development, we would probably goof as often and as badly as run-of-the-mill parents.

I should like, first, to remind you of three of the most basic and general of the propositions in that theory of motivation which has been dominant for the past thirty to forty years. These are propositions which, although stated in somewhat varied forms, have been shared by both psychoanalysts and academic behaviour theorists. Secondly, I wish to cite evidence which calls these propositions into question, and thirdly, to suggest, tentatively, three new interpretative principles which appear to me to be congruent with a large number of facts and which have interesting implications.

Our conceptions of motivation have traditionally been concerned with three large questions: (*a*) Why does an organism or person become active?, (*b*) Why does the organism or person act one way rather than another? and (*c*) How do you get the organism or person to change his behaviour to something conceived to be more desirable or appropriate?

The dominant theory

Drive

According to our dominant theory, it is claimed, first of all, that 'all

---

Source: *Child Development*, 1960, *31*, 489–504.

behaviour is motivated', and that the aim or function of every instinct, defence, action, or habit is to reduce or eliminate stimulation or excitation within the nervous system. It is not easy to state when this view was first presented. Signs of it appear in the seventh chapter of Freud's *Interpretation of Dreams* in 1900, and the idea is full-blown in his paper entitled 'Instincts and their Vicissitudes' in 1915. The idea also appears in Woodworth's *Dynamic Psychology* (1918), where the term *drive* was first introduced into the glossary of American psychology. The idea was full-blown in Dashiell's *Fundamentals of Objective Psychology* (1928).

Although Freud (1915) believed that the source of motivation lay outside the domain of psychology in physiology, American psychologists, untroubled by such limits to their domain, have gone on to answer the first question concerning what motivates organisms to become active by saying that they are *driven.* Organisms have been conceived to be driven, first by those so-called primary inner stimuli which arise from homoeostatic imbalances or needs. With no shame whatsoever, psychologists have long cited the evidence from the work of such physiologists as Claude Bernard (1859) and his successors, and especially of Walter B. Cannon (1915), and also of the psychologist Curt Richter (1927) to document this answer. Organisms are driven, second, by various forms of intense and painful external stimulation. It has been assumed that these two forms of stimulation arouse an inner state of excitement which has usually been called *drive.*

It is also assumed, as the proposition that 'all behaviour is motivated' implies, that the organism would be inactive unless driven by either inner or outer stimuli. Freud (1915) has been highly explicit about this assumption, and the assumption lies implicitly behind the notion of conditioned or learned drive in behaviour theory and behind the traumatic notion of anxiety in psychoanalysis. It is obvious, of course, that animals and people are sometimes active when it is hard to see how either homoeostatic drive or painful external stimulation could be operative. It is then assumed that some of the weak, innocuous stimuli present must have been associated in the past with either painful stimuli or homoeostatic needs. In such a way the weak stimuli which are present must have acquired the capacity to arouse the drive, often now called anxiety by psychologists as well as psychoanalysts, and it is such acquired or conditioned drive that is conceived to activate the organism.

Such conditioned drive or anxiety has been well demonstrated in the laboratory. Before World War II, Miller (1941, 1948) at Yale showed that rats which had been repeatedly shocked in a white box would, when later returned to the white box, make an effort to escape. Moreover, in the course of these efforts, they could be got to learn new skills such as that of turning a wheel to open a door. Rats which had not been shocked in the white box made no such efforts to escape. In another demonstration, Solomon and Wynne (1953) have shown that dogs which have experienced a tone or a buzzer paired a few times with a subtetanizing shock will run away from that tone or buzzer for hundreds of trials, with the average reaction time of starting continuing to decrease through 600 such trials. In my own work (Hunt, 1941) rats fed irregularly in infancy ate more than their litter-mate controls and sometimes (Hunt 1947) hoarded food in adulthood after a period without food. Here, as I conceived it, the cues of hunger were conditioned to intense hunger excitement during the infantile experience. In adulthood the conditioned hunger drive facilitated the rate of eating and, sometimes, hoarding.

Such work has demonstrated that this notion of conditioned drive or anxiety, which goes back to the work of Bechterev (1913) and Watson and Raynor (1920), has a solid basis in reality. But in what has been the dominant theory of motivation, as epitomized by Freud's (1926) later traumatic theory of anxiety, and by the Hull (1943) and Dollard–Miller theory of acquired drives (Miller and Dollard, 1941; Dollard and Miller, 1950), conditioning is conceived to be the only way in which an organism can become fearful of innocuous stimuli.

Habit

Habit has been the answer to the second question concerned with why an animal or person acts one way rather than another. The organism is controlled by the habits which have served to reduce drive in the past when that organism was in the presence of the inner and outer drive stimuli, and the cue stimuli impinging upon him at any given moment now. Under the term *habit*, I am including psychoanalytic modes, which have supposedly been fixated during infancy in the course of either too much gratification or too much frustration, and I am including also ego-defences, or anxiety equivalents, and cathexes, as well as the instrumental responses and traits commonly investigated in psychological laboratories.

Changing behaviour has been conceived to be a matter of motivating the organism with either punishment or homoeostatic need to make the desired behaviour which can then be reinforced by arranging for it to reduce the drive aroused by the punishment or the need. Although the conditions and conceptions of psychotherapy in the clinic differ considerably from the conditions and conceptions of the behaviour theorist investigating learning in laboratory animals, in either case it is conceived that motivation is a necessity, and motivation means changing the emotional or drive conditions which are quite extrinsic to either the instrumental behaviour or the cognitive, informational processes concerned.

This dominant theory has been a conceptual edifice of large dimensions and of considerable detail. It has provided a plausible account of both personality development and social motives. The experimental facts of homoeostasis and of conditioned drive and fear are sound. Nevertheless, it has become more and more evident in the past ten years that some of the basic assumptions of this dominant theoretical scheme and some of the explanatory extrapolations contradict facts and call for reinterpretation.

Reinterpretations

Is all behaviour motivated?

The first of the assumptions to be called into question is the one that *all behaviour is motivated* and that *organisms become inactive unless stimulated* by homoeostatic need or painful stimulation or conditional stimuli for these. A large variety of observations contradict this assumption and imply spontaneous molar activity. Beach (1945) has reviewed the observations of play in the young to show that playful activities are most likely to occur when either young animals or children are homoeostatically satisfied and also comfortably warm. The very occurrence of either homoeostatic need or strong external stimulation stops play and turns the young animal or child to activities calculated to relieve such stimulation. Berlyne (1950, 1955) has shown that well-fed and watered rats will explore areas new to them if given only the opportunity. Montgomery (1953) moreover, has shown that hunger and thirst tend to limit the exploratory behaviour of rats rather than facilitate it, and Montgomery and Monkman (1955), as well as others, have shown that conditioned fear inhibits exploration. Harlow, Harlow, and Meyer (1950) have demonstrated that well-fed monkeys will learn to unassemble a three-device puzzle with no other drive and 'no other reward than the privilege of unassembling it'. In another study Harlow (1950) found two well-fed and well-watered monkeys worked repeatedly at unassembling a six-device puzzle for ten continuous hours, and they were still showing what he characterized as enthusiasm for their work on the tenth hour of testing. From his observations of the human child, moreover, Piaget (1952) remarks repeatedly on the enthusiastic and repeated performance of such emerging skills as the release of a toy, sitting up, standing, etc.

Such evidences of spontaneous behaviour, which is unmotivated in the traditional sense, have led to the naming of such new motives as a curiosity drive by Berlyne (1955), an exploratory drive by Montgomery (1951), and exteroceptive and curiosity drives by Harlow (1953). I would like to object that merely naming such drives explains nothing. If we continue we shall be revisiting McDougall's (1915) practice of postulating a separate drive for almost every variety of activity. Let us stop with noting that such observations do contradict our assumption that organisms will become inactive unless driven by homoeostatic needs and painful stimuli, and give up this ancient Greek notion that living matter is inert substance to which motion must be imparted by extrinsic forces. We can then embrace the thermodynamic conception of living things as open systems of energy exchange which exhibit activity intrinsically and upon which stimuli have a modulating effect, but not an initiating effect.

This notion of activity being intrinsic in living tissue is receiving support from studies of organ systems as well as studies of molar organisms. The EEG, for example, shows that brain cells are continuously active (Jasper, 1937; Prosser, 1934). In sleep the slow waves of large amplitude are taken to imply that large numbers of cells are firing synchronously, and the effect of waking and stimulation and exciting the brain-stem-reticular formation is to asynchronize this firing which shows in rapid waves of low magnitude (Lindsley, 1957).

Granit (1955) points out that the spontaneous firing of retinal cells increases with dark adaptation and thereby functions to prevent the de-afferentization of visual cortex with darkness. Twenty years ago, this spontaneous firing was considered, at worst, to be due to some failure of experimental control, or at best, noise in the channel of information. Recently, the Laceys (1958) have found spontaneous fluctuations of sudomotor activity and cardiac activity which they see as functioning in the control of the organism's relations with its environment. Especially intriguing is their notion that the carotid sinus mechanism functions as a feedback loop which participates in the directing of attention inward or outward by inhibiting or facilitating receptor inputs. But the point of mentioning these evidences of spontaneous activities of organ systems here is merely to help inter for good the notion that activity of living systems requires homoeostatic need or painful external stimulation and to foster the idea that to live means to be active in some degree.

Reinforcement

This idea of activity being intrinsic in living organisms has implications for our conception of reinforcement. It makes it unnecessary to see an activity as a matter of either reducing or avoiding stimulation which is implied in the assumption that organisms become inactive unless stimulated. This is a second fundamental assumption of the dominant theory which has been shared by psychoanalysts and behaviour theorists alike.

On the one hand, there is still a place for drive reduction. It is clear that under conditions of homoeostatic need and painful stimulation, and perhaps under circumstances when the conditions of stimulation are changing with too great rapidity, both animals and persons learn techniques and strategies leading to gratification or reduction in external stimulation. The evidence that led Thorndike to formulate the 'law of effect' is as convincing as ever. Moreover, in association with reductions of homoeostatic needs, animals and men may also learn cathexes or emotional attachments. The facts referred to are those highly familiar in secondary reinforcement (Hull, 1943; Pavlov, 1927).

On the other hand, the facts implying that organisms show spontaneous molar activity also imply that, when animals and human beings have been living under conditions of low and unchanging stimulation for a time, increases of stimulation become reinforcing. Butler has shown that rhesus monkeys will learn quite complex discriminations with the only reward being a peek through a glass window (Butler, 1953) at the things in the next room or a few seconds of auditory experience (Butler, 1957). Berlyne (1950) has shown that, the greater the variety of stimulation in an area which rats are permitted to explore, the longer they continue their explorations.

Especially important in this connection are the studies of human behaviour under conditions of minimal variation in stimulation. I refer to the studies of perceptual isolation by Bexton, Heron, and Scott (1954) at McGill and also the work of Lilly (1956). At McGill, college students were paid 20 dollars a day to do nothing. They lay for twenty-four hours a day on a comfortable bed. The temperature was optimal and constant. Eyes, ears, and hands were shielded to minimize stimulus variation. Few subjects could endure more than two or three days of such conditions. They developed a desire for variation which was almost overwhelming.

While interpreting such facts in terms of a multiple set of drives for curiosity, exploration, or stimulation will get us only to a redescription of them, Hebb's (1955) notion of an optimal level of activation–and, I would like to add, stimulus variation below which *increases* are reinforcing and above which *decreases* are reinforcing–is an integrative conception of fair magnitude. Moreover, the drive-reduction principle of reinforcement may be seen to be but half of this more general curvilinear principle.

But this is probably not the whole story. It looks as if there were, natively, both positive and negative forms of exciting stimulation. Sheffield, Roby, and Campbell (1954) have argued that the reinforcing effect of eating is not a matter of reduction of the hunger drive but rather a matter of the positive value of the consummatory act of eating. Moreover, Sheffield, Wulff, and Backer (1951) have shown that male rats will learn mazes to get to females in heat even when they are allowed only intromission but not allowed to continue coitus to the point of drive-reducing ejaculation. From the fact that Davis and Buchwald (1957) at Indiana have shown that showing pictures of nude women to college males increases excitement as shown by increased palmar conductance and the arrest of EEG-alpha, it is clear that such stimulation is exciting rather than excitement-reducing. Young (1955) has long emphasized the importance of the hedonic quality of experience for reinforcement, and he has shown that speed of running in rat subjects increases with the concentration of sucrose in the incentive drink.

The suggestion that the two forms of excitation, one positive and one negative, are built into organisms comes also from the work of Olds and Milner (1954). Electrical stimulation of the septal area is positively reinforcing, but electrical stimulation of the brain-stem reticular formation is negatively reinforcing. Perhaps, it is not without significance that the septal area is part of the old olfactory brain which has been considered to have an especially important part in the mediation of sexual and consummatory behaviour in mammals. At any rate, it looks as though certain types of stimulation may be positively reinforcing even though they be intense and exciting. This may mean that the curvilinear principle may be limited in its domain to strong stimulation via the exteroceptors when homoeostatic needs are minimized.

The suggestion of innate, positive and negative exteroceptive stimulation comes secondly from recent work by Harlow (1958). It has been customary to see an infant's cathexes or love for its mother developing as secondary reinforcement largely out of its feeding experiences. Freud (1905), of course, contended that the pleasure from stimulation of the oral erogenous zone furnished the experiential basis for both pleasure-sucking and maternal attachment, a contention which contradicted his most definitive formulations of drive theory (Freud, 1915). The fact that an infant must suck for its nourishment, according to libido theory (Freud, 1905, p. 587), merely guaranteed discovery of the pleasures of oral stimulation. Behaviour theorists have seen both sucking and love of mother as forms of secondary reinforcement deriving from the fact that the child satisfies its hunger by means of sucking the mother's breasts (Mussen and Conger, 1956, pp. 137ff.). Harlow (1958), however, has recently compared the degree of attachment of young monkeys to a wire mother-surrogate on which they nursed at a bottle with attachment to a padded and cloth-covered mother-surrogate on which they received nothing but the feel of softness. In terms of the amount of time spent on each of the two mother-surrogates, the monkeys showed more than ten times the attachment to the soft-padded surrogate as to the wire surrogate. When various fear-evoking stimuli were presented to the baby monkeys in their cages, it was to the padded and cloth-covered surrogate that the frightened infant monkey turned, not to the wire surrogate on which it had been nursed. Harlow argues from these findings that it is the sensory quality of softness which gives the reinforcement. His study suggests, moreover, that it is important to investigate the capacity for various kinds of stimuli for positive and negative reinforcement in the very young. Pratt (1954) cites a monograph by Canestrini (1913) on the sensory life of the newborn for an observation that certain stimuli are associated with decreases in the rate of the heart rate, and are therefore pleasant, while others are associated with increases in heart rate and are unpleasant.[1] In view of the finding by Davis and Buchwald

---

[1] An examination of Canestrini's monograph shows that Pratt was mistaken in stating that Canestrini remarked upon decreases in heart rate being associated with pleasure, but some of his published kymograph records do indicate decreases in heart rate. It may well be that heart rate could serve as an indicator of the emotional value of various sensory inputs and these might be tested for their reinforcement values. I am indebted to Dr William Gerler for reading this monograph carefully to check my own impression of Canestrini's text.

(1957) that seeing a picture of a nude female results in reduction of the heart rate of male college students, it is possible that this physiological indicator may provide a technique for determining the direction of the reinforcing effect of stimuli in the newborn. At any rate, what is suggested is that McDougall's (1915) old notion of natively positive and negative values for receptors inputs be re-examined.

Conditioned fear and anxiety

The third assumption that I wish to examine in the light of empirical evidence is the notion that fear and anxiety are *always* inculcated as a consequence of traumatic experiences of helplessness in the face of homoeostatic need or painful external stimulation. Note that I am not denying that such conditioned fears do exist, I am only questioning the word *always* . . . are always inculcated as a consequence of traumatic experiences.

The first relevant studies go way back to the 1920s. Harold and Mary Cover Jones (1928) attempted to test the claims of Watson (1928) and Watson and Raynor (1920) concerning conditioned fears. They exposed their subjects of various ages, ranging from early infancy to adult, to a large but sluggish and harmless bull-snake. Fear of the snake was exceedingly common among adults, teenagers, and latency-age children, but it was absent in children below three years of age. It began to appear among children older than three and was typical of children six and older. From the fact that the fear appeared at a younger age in those of higher intelligence than those of lower intelligence, the Joneses argued that fear of snakes is a response which comes automatically into the developing child's repertoire through maturation. This remains as an alternative hypothesis to that of conditioned fear.

A study by Frances Holmes (1935), which is seldom cited, calls both of these interpretations into questions. Holmes compared the fearfulness of the children of lower-class background who were attending a day nursery, with the fearfulness of children of upper-class background who were attending a private nursery school. She got her fear scores by indicating that the child could get some attractive toys with which to play by going into the dark room adjacent to the examining room, or by taking them off a chair situated beside that of a strange woman dressed in a large floppy black hat and a long grey coat, or by climbing along a plank some three feet off the floor. If the child started immediately for the toys, he got a score of one for that item. If he hesitated but ultimately went ahead on his own, he got a score of two. If he would go only if accompanied by the examiner, the score was three. If he refused to go at all, the score was four. There were seven such situations. The results show that the fear scores of the lower-class children averaged only about half the size of those for the upper-class children, and the fear scores for boys were lower than those for girls. Yet it would be the lower-class children who had experienced the more homoeostatic need and painfully rough treatment than the upper-class children, and the boys had probably experienced more painful experiences than the little girls. That intelligence is not the factor is shown by the fact that the fear scores showed a correlation of only about +0.2 with mental age, and the differences were still significant when intelligence was partialled out. Something besides either conditioned fear or the correlation between fear and intelligence is required to make these results comprehensible.

Recently, evidence even more contradictory to the notion of conditioned fears has been coming from the work of Seymour Levine. Levine, Chevalier, and Korchin (1956) have compared the adult behaviour of rats shocked and rats petted daily from birth to the twentieth day with the adult behaviour of rats left continuously in their nest with their mothers. When he started his work, Levine expected to find that the shocked animals would show traumatic effects of their shock experiences in heightened emotionality and damaged capacity to learn adaptive responses. On the contrary, the shocked animals, along with the handled animals gained weight faster than those left in the nest (Levine, 1957a, 1957b, 1958; Levine *et al.*, 1956). Byron Lindholm, working with the writer, has repeated and confirmed this finding. Moreover, Levine's shocked and handled animals both showed less emotionality than those left continuously in the nest with their mothers, i.e. less emotionality in the sense that they defecated and urinated less frequently when placed in a strange situation. Finally, the shocked and handled animals, which have appeared alike in all of these experiments, learned an avoidance response more rapidly and drank more readily after eighteen hours without water than did the rats left in the nest with their mother.

Clearly these results on both human children and rats imply that fear and anxiety must sometimes have some other basis than that of being associated with painful stimulation. Hebb (1946, 1949) has formulated a radically different explanation of fear which may be termed either an incongruity or a dissonance theory.

The facts which suggested Hebb's conception came largely from observing chimpanzees being raised under controlled conditions at the Yerkes Laboratory. Fear, defined as withdrawal behaviour in response to the appearance of some object, does not appear in young chimpanzees until they are approximately four months old. Then, the objects feared are familiar objects in unfamiliar guise. Fear of strangers is an example. This appears spontaneously to the first stranger seen, so it cannot be based on associating strangers with painful stimulation. Fear of strangers does not appear in chimpanzees—or in children, I might add—who have always been exposed to a large number of persons. While the avoidance response is unlearned, the familiar, expected aspects of objects must be learned. The young animal must have established as residues of his experience cortical firing patterns (or cognitive structures—whichever term you like) from which new receptor inputs can be incongruous. Consider the kinds of objects regularly feared. They are, for instance, the familiar keeper or experimenter in strange clothes, the experimenter in a Halloween mask, a plaster cast of a chimpanzee head (which lacks, of course, the familiarly attached body), an anaesthetized chimpanzee infant (from which the familiar patterns of motion are absent). On the other hand, objects which have never entered into the young chimpanzee's life may be strange without evoking withdrawal. In other words, the feared object is one which excites receptors in a fashion which is incongruous with the central, sequential pattern of neural firing which has accrued as a residue of the chimpanzee or human infant's past experience. Until the central pattern has been learned, incongruous stimulation is impossible.

Such a conception can well account for Holmes' findings that lower-class children are less fearful than higher-class children and that boys are less fearful than girls even though both lower-class children and boys of nursery school age are likely to have had the wider experience with the sorts of situations used by Holmes to evoke fear. It may well be that being shocked and handled provides a variety of experience which leaves the rat pups which have been subjected to it less disturbed by such things as open fields and eighteen hours without water, but these effects may ultimately be found to be a matter of still another mechanism. It is too early to say.

Taking seriously this incongruity-dissonance conception of the genesis of fear leads to interesting reinterpretations of a great many of the motivational phenomena of child development. Consider these few. In considering separation anxiety, the incongruity principle makes it necessary to puzzle about how the absence of mother could be the conditional stimulus for the traumatizing and helpless distress that has been supposed to have occurred in her absence. In considering fear of the dark, it also becomes unnecessary to puzzle about how the absence of light stimulation could so widely have been associated with painful stimulation. Multiple mothering need not be seen as a traumatizing experience in the light of this conception, but rather as an innoculation against social shyness and fear. The timidity of the overprotected child and the social shyness of the rural mountain people get an explanation which has been difficult in terms of the theory of conditioned fear.

Motivation in terms of the incongruity-dissonance principle

This introduction of the incongruity-dissonance principle concludes the three reinterpretations I wish to present, but I do wish to call your attention to the pervasive character of this incongruity-dissonance principle. It appears to have great explanation power which figures, in one guise or another, in several systematic theories besides that of Hebb, all of which have been characterized as non-dynamic.

Hebb's (1949) theorizing is physiological, at least in a verbal sense, in that he conceives the residues of past inputs to be stored in semi-autonomous, reverberating cerebral circuits which he terms *cell assemblies.* These cell assemblies are the neural analogue of concepts, and they get sequentially integrated into what he calls *phase sequences.* The sequential organization in time provides for the subjective phenomenon of expectation. When markedly incongruous receptor inputs disrupt this sequential organization, behaviour is changed and the process is felt as unpleasant emotion. Slight degrees of incongruity, which can readily be accommodated, lend interest and may provide attractive problems, but the larger ones are repelling and perhaps even devastating.

Piaget (1952, 1954) utilizes very much the same incongruity notion to account for the development of intelligence and concepts in human children. In his system, the child comes at birth with certain sensory-motor coordinations which he terms *schemata.* Variation in stimulus situations calls for adaptive *accommodations* or changes in these schemata, which changes are *assimilated* or stored as residues. Piaget also finds limited incongruities between central schemata and receptor inputs to be interesting and facilitative of growth, but incongruities which extend beyond the child's capacity for accommodation instigate withdrawal or fear and even terror. In Piaget's theory the child's gestalt-like conceptions of reality (space, time, and number) are schemata which develop through a continuous process of accommodations and assimilations and become fixed or static only when the child's schemata come to correspond so well with reality that no further accommodations are required. Here agreement among people is dictated by reality.

Helson (1947, 1948) has called the residues of immediate past experience in the typical psychophysical experiment an *adaptation level.* Both he and McClelland (1953) have seen affective arousal to be a matter of the size of the discrepancy between receptor inputs and the adaptation level. Small discrepancies may be attractively pleasant, large ones repellingly unpleasant. As an example, some of you will readily recall having experienced the affective startle that comes when you have been set to pick up what you thought was a full pail, only to find it empty.

Festinger (1957) has recently written a book entitled *A Theory of Cognitive Dissonance* in which he shows that a discrepancy between belief about a situation and perception of that situation acts like a drive. The subject acts to reduce the *dissonance* by either withdrawing from the incredible situation or by changing his beliefs, and, not incidentally, he finds the dissonance highly unpleasant.

Rogers (1951) has described the basis for anxiety as discrepancy between the 'phenomenological field' and the perceived reality as represented by his two circles. Rogers' phenomenological field, however, is not the perceptually-given phenomenal field of such German phenomenologists as Delthei and Husserl. It is rather the inferred storehouse of past experience and represented in the present by expectations, aspirations, self-concept, and the like. Thus, his conceptual scheme appears to fall within the domain of the incongruity-dissonance principle.

Kelly's (1955) *Psychology of Personal Constructs* also makes central use of this principle. The term *personal constructs* refers to the ways in which individuals construe and anticipate events. These each person derives from the way in which he has experienced such events in the past. When a person's constructions fail to predict events, this is disturbing, even anxiety-producing, and it motivates some kind of change but the change may take place in defences against such change of constructs or in avoiding such events, or in the constructs themselves.

Perhaps, it is worth noting in closing that this incongruity-dissonance principle makes both motivation and reinforcement intrinsic to the organism's relations with its environment, intrinsic, if you will, to the organism's information-processing. It is as if the organism operated like an error-actuated, feedback system where the error is derived from discrepancy between receptor-inputs of the present and the residues of past experience which serve as the basis for anticipating the future. The dominant view of the past half century has seen both motivation and reinforcement as extrinsic to the information-processing. This has put a tremendous burden of responsibility for the management of affective motivation on parents, teachers, and all those in positions of authority and control. Visions of man completely controlled as exemplified by George Orwell's *1984*, are conceivable only by assuming that the extrinsic motivating forces of homoeostatic need and painful stimulation are completely dominant. In this light the terror of the baby chimp at seeing his keeper in a Halloween mask and the irritation of the believer when his beliefs are disconfirmed are perhaps symbols of hope. They may justify Abraham Lincoln's well-known dictum that 'you can fool some of the people all the time, and all the people some of the time, but you cannot fool all the people all the time'.

To return to the cartoon of the lad who was improperly motivated: Perhaps, the task of developing proper motivation is best seen, at least in nutshell form, as limiting the manipulation of extrinsic factors to that minimum of keeping homoeostatic need and exteroceptive drive low, in favour of facilitating basic information-processing to maximize accurate anticipation of reality.

References

BEACH, F. A. (1945) 'Current concepts of play in animals', *Am. Naturalist, 79*, 523–41.

BECHTEREV, V. M. (1913) *La Psychologie objective,* trans. N. KOSTYLEFF. Paris: Alcan.

BERLYNE, D. E. (1950) 'Novelty and curiosity as determinants of exploratory behavior', *Br. J. Psychol., 41*, 68–80.

BERLYNE, D. E. (1955) 'The arousal and satiation of perceptual curiosity in the rat', *J. comp. physiol. Psychol., 48,* 238–46.

BERNARD, C. (1859) *Leçons sur les propriétés physiologiques et les alternations pathologiques des liquides de l'organisme.* Paris: Ballière, 2 vols.

BEXTON, W. H., HERON, W., and SCOTT, T. H. (1954) 'Effects of decreased variation in the sensory environment', *Canad. J. Psychol., 8,* 70–6.

BUTLER, R. A. (1953) 'Discrimination learning by rhesus monkeys to visualexploration motivation', *J. comp. physiol. Psychol., 46,* 95–8.

BUTLER, R. A. (1957) 'Discrimination learning by rhesus monkeys to auditory incentives', *J. comp. physiol. Psychol., 50,* 239–41.

CANESTRINI, S. (1913) 'Uber das Sinnesleben des Neugebornen', in A. ALZHEIMER and M. LEWANDOWSKY (eds.) *Monogr. Gesamt. Neurol. Psychiat.* (Heft 5), Berlin: Springer.

CANNON, W. B. (1915) *Bodily Changes in Pain, Hunger, Fear, and Rage.* New York: Appleton-Century.

DASHIELL, J. (1928) *Fundamentals of Objective Psychology.* New York: Houghton Mifflin.

DAVIS, R. C., and BUCHWALD, A. M. (1957) 'An exploration of somatic response patterns: stimulus and sex differences', *J. comp. physiol. Psychol., 50,* 44–52.

DOLLARD, J., and MILLER, N. E. (1950) *Personality and Psychotherapy.* New York: McGraw-Hill.

FESTINGER, L. A Theory of Cognitive Dissonance. *New York: Row, Peterson.*

FREUD, S. (1900) 'The interpretation of dreams', in *The Basic Writings of Sigmund Freud,* trans. A. A. BRILL. New York: Modern Library, 1938, pp. 179–548.

FREUD, S. (1905) 'Three contributions to the theory of sex', in *The Basic Writings of Sigmund Freud,* trans. A. A. BRILL. New York: Modern Library, 1938, pp. 553–629.

FREUD, S. (1915) 'Instincts and their vicissitudes', in *Collected Papers,* vol. iv. London: Hogarth Press, 1950, pp. 60–83.

FREUD, S. (1926) *Inhibition, Symptom, and Anxiety,* trans. H. A. BUNKER as *The Problem of Anxiety.* New York: Norton, 1936.

GRANIT, R. (1955) *Receptors and Sensory Perception.* Yale University Press.

HARLOW, H. F. (1950) 'Learning and satiation of response in intrinsically motivated complex puzzle performance by monkeys', *J. comp. physiol. Psychol., 43,* 289–94.

HARLOW, H. F. (1953) 'Motivation as a factor in the acquisition of new responses.' In *Current Theory and Research in Motivation: a symposium,* University of Nebraska Press, pp. 24–49.

HARLOW, H. F. (1958) 'The nature of love',-*Am. Psychol., 13,* 673–85.

HARLOW, H. F., HARLOW, M. K., and MEYER, D. R. (1950) 'Learning motivated by a manipulation drive', *J. exp. Psychol., 40,* 228–34.

HEBB, D. O. (1946) 'On the nature of fear', *Psychol. Rev., 53,* 259–76.

HEBB, D. O. (1955) 'Drives and the CNS (conceptual nervous system)', *Psychol. Rev., 62,* 243–54.

HELSON, H. (1947) 'Adaptation-level as frame of reference for prediction of psycho-physical data', *Am. J. Psychol., 60,* 1–29.

HELSON, H. (1948) 'Adaptation-level as a basis for a quantitative theory of frames of reference', *Psychol. Rev., 55,* 297–313.

HOLMES, F. B. (1935) 'An experimental study of the fears of young children', in A. T. JERSILD and F. B. HOLMES, 'Children's fear', *Child Developm. Monogr., 20,* 167–296.

HULL, C. L. (1943) *Principles of Behavior.* New York: Appleton-Century.

HUNT, J. MCV. (1941) 'The effects of infant feeding-frustration upon adult hoarding in the albino rat', *J. abnorm. soc. Psychol., 36,* 338–60.

HUNT, J. MCV., SCHLOSBERG, H., SOLOMON, R. L., and STELLAR, E. (1947) 'Studies on the effects of infantile experience on adult behavior in rats: I. Effects of infantile feeding frustration on adult hoarding', *J. comp. physiol. Psychol., 40,* 291–304.

JASPER, H. H. (1937) 'Electrical signs of cortical activity', *Psychol. Bull., 34,* 411–81.

JONES, H. E., and JONES, M. C. (1928) 'A study of fear', *Child Educ., 5,* 136–43.

KELLY, G. A. (1955) *The Psychology of Personal Constructs.* New York: Norton.

LACEY, J. I., and LACEY, B. C. (1958) 'The relationship of resting autonomic activity to motor impulsivity', in *The Brain and Human Behavior.* Baltimore: Williams & Wilkins, pp. 144–209.

LEVINE, S. (1957a) 'Infantile experience and consummatory behavior in adulthood', *J. comp. physiol. Psychol., 50,* 609–12.

LEVINE, S. (1957b) 'Infantile experience and resistance to physical stress', *Science, 126,* 405.

LEVINE, S. (1958) 'Noxious stimulation in infant and adult rats and consummatory behavior', *J. comp. physiol. Psychol., 51,* 230–3.

LEVINE, S., CHEVALIER, J. A., and KORCHIN, S. J. (1956) 'The effects of shock and handling in infancy on later avoidance learning', *J. Personality, 24,* 475–93.

LILLY, J. C. (1956) 'Mental effects of reduction of ordinary levels of physical stimuli on intact, healthy persons', *Psychiat. Res. Rep., 5,* 1–9.

LINDSLEY, D. B. (1957) 'Psychophysiology and motivation', in M. R. JONES (ed.) *Nebraska Symposium on Motivation.* University of Nebraska Press, pp. 44–105.

MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., and LOWELL, E. L. (1953) *The Achievement Motive.* New York: Appleton-Century-Crofts.

MCDOUGALL, W. (1915) *An Introduction to Social Psychology.* Boston: Luce.

MILLER, N. E. (1941) 'An experimental investigation of acquired drives', *Psychol. Bull., 38,* 534–5.

MILLER, N. E. (1948) 'Studies of fear as an acquirable drive: I. Fear as motivation and fear-reduction as reinforcement in the learning of new responses', *J. exp. Psychol., 38,* 89–101.

MILLER, N. E., and DOLLARD, J. (1941) *Social Learning and Imitation.* Yale University Press.

MONTGOMERY, K. C. (1951) 'The relation between exploratory behavior and spontaneous alternation in the white rat', *J. comp. physiol. Psychol., 44,* 582–9.

MONTGOMERY, K. C. (1953) 'The effect of the hunger and thirst drives upon exploratory behavior', *J. comp. physiol. Psychol., 46,* 315–19.

MONTGOMERY, K. C., and MONKMAN, J. A. (1955) 'The relation between fear and exploratory behavior', *J. comp. physiol. Psychol., 48,* 132–6.

MUSSEN, P. H., and CONGER, J. J. (1956) *Child Development and Personality.* New York: Harper.

OLDS, J. (1955) 'Physiological mechanisms of reward', in M. R. JONES (ed.) *Nebraska Symposium on Motivation,* University of Nebraska Press, pp. 73–139.

OLDS, J., and MILNER, P. (1954) 'Positive reinforcement produced by electrical stimulation of septal area and other regions of the rat brain', *J. comp. physiol. Psychol., 47,* 419–27.

PAVLOV, I. P. (1927) *Conditioned Reflexes,* trans. C. V. ANREP. New York: Oxford University Press.

PIAGET, J. (1952) *The Origins of Intelligence in Children.* New York: International Universities Press.

PIAGET, J. (1954) *The Construction of Reality in the Child,* trans. M. COOK. New York: Basic Books.

PRATT, K. C. (1954) 'The neonate', in L. CARMICHAEL (ed.) *Manual of Child Psychology,* 2nd edn. New York: Wiley, pp. 215–91.

PROSSER, C. L. (1934) 'Action potentials in the nervous system of the crayfish: I. Spontaneous impulses', *J. cell. comp. Physiol., 4,* 185–209.

RICHTER, C. P. (1927) 'Animal behavior and internal drives', *Quart Rev. Biol., 2,* 307–43.

ROGERS, C. R. (1951) *Client-centred Therapy.* New York: Houghton Mifflin.

SHEFFIELD, F. D., ROBY, T. B., and CAMPBELL, B. A. (1954) 'Drive reduction versus consummatory behavior as determinants of reinforcement', *J. comp. physiol. Psychol., 47,* 349–55.

SHEFFIELD, F. D., WULFF, J. J., and BACKER, R. (1951) 'Reward value of copulation without sex drive reduction', *J. comp. physiol. Psychol., 44,* 3–8.

SOLOMON, R. L., and BRUSH, E. S. (1956) 'Experimentally derived conceptions of anxiety and aversion', in M. R. JONES (ed.) *Nebraska Symposium on Motivation,* University of Nebraska Press, pp. 212–305.

SOLOMON, R. L., and WYNNE, L. C. (1953) 'Traumatic avoidance learning: acquisition in normal dogs', *Psychol. Monogr., 67,* no. 4 (whole no. 354).

THORNDIKE, E. L. (1913) *Educational Psychology.* Vol. i, *The original nature of man,* vol. ii, *The Psychology of Learning.* New York: Teachers College.

WATSON, J. B. (1928) *Psychological Care of the Infant and Child.* New York: Norton.

WATSON, J. B., and RAYNOR, R. (1920) 'Conditional reactions', *J. exp. Psychol., 3,* 1–4.

WOODWORTH, R. S. (1918) *Dynamic Psychology*. Columbia University Press.

YOUNG, P. T. (1955) 'The role of hedonic processes in motivation', in M. R. JONES (ed.) *Nebraska Symposium on Motivation*, University of Nebraska Press, pp. 193–237.

11 The will to learn

J. S. Bruner

The single most characteristic thing about human beings is that they learn. Learning is so deeply ingrained in man that it is almost involuntary, and thoughtful students of human behavior have even speculated that our specialization as a species is a specialization for learning. For, by comparison with organisms lower in the animal kingdom, we are ill equipped with prepared reflex mechanisms. As William James put it decades ago, even our instinctive behavior occurs only once, thereafter being modified by experience. With a half century's perspective on the discoveries of Pavlov, we know that man not only is conditioned by his environment, but may be so conditioned even against his will.

Why then invoke the idea of a 'will to learn'? The answer derives from the idea of education, a human invention that takes a learner beyond 'mere' learning. Other species begin their learning afresh each generation, but man is born into a culture that has as one of its principal functions the conservation and transmission of past learning. Given man's physical characteristics, indeed, it would be not only wasteful but probably fatal for him to reinvent even the limited range of technique and knowledge required for such a species to survive in the temperate zone. This means that man cannot depend upon a casual process of learning; he must be 'educated'. The young human must regulate his learning and his attention by reference to external requirements. He must eschew what is vividly right under his nose for what is dimly in a future that is often incomprehensible to him. And he must do so in a strange setting where words and diagrams and other abstractions suddenly become very important. School demands an orderliness and neatness beyond what the child has known before; it requires restraint and immobility never asked of him before; and often it puts him in a spot where he does not *know* whether he knows and can get no indication from anybody for minutes at a time as to whether he is on the right track. Perhaps most important of all, school is away from home with all that fact implies in anxiety, or challenge, or relief.

In consequence of all this the problem of 'the will to learn' becomes important, indeed exaggerated. Let us not delude ourselves: it is a problem that cannot be avoided, though it can be made manageable, I think. We shall explore what kinds of factors lead to satisfaction in 'educated' learning, to pleasure in the practice of learning as it exists in the necessarily artificial atmosphere of the school. Almost all children possess what have come to be

---

Source: BRUNER, J. S. (1966) *Toward a Theory of Instruction.* Cambridge, Mass.: Harvard University Press, 113–28.

called 'intrinsic' motives for learning. An intrinsic motive is one that does not depend upon reward that lies outside the activity it impels. Reward inheres in the successful termination of that activity or even in the activity itself.

Curiosity is almost a prototype of the intrinsic motive. Our attention is attracted to something that is unclear, unfinished, or uncertain. We sustain our attention until the matter in hand becomes clear, finished, or certain. The achievement of clarity or merely the search for it is what satisfies. We would think it preposterous if somebody thought to reward us with praise or profit for having satisfied our curiosity. However pleasant such external reward might be, and however much we might come to depend upon it, the external reward is something added. What activates and satisfies curiosity is something inherent in the cycle of activity by which we express curiosity. Surely such activity is biologically relevant, for curiosity is essential to the survival not only of the individual but of the species. There is considerable research that indicates the extent to which even nonhuman primates will put forth effort for a chance to encounter something novel on which to exercise curiosity. But it is clear that unbridled curiosity is little more than unlimited distractibility. To be interested in everything that comes along is to be interested in nothing for long. Studies of the behavior of three-year-olds, for example, indicate the degree to which they are dominated from the outside by the parade of vivid impressions that pass their way. They turn to this bright color, that sharp sound, that new shiny surface. Many ends are beyond their reach, for they cannot sustain a steady course when the winds shift. If anything, they are 'too curious'. They live by what psychologists have long called the laws of primary attention: attention dominated by vividness and change in the environment. There has been much speculation about the function of this early and exhausting tempo of curiosity. One neuropsychologist, Donald Hebb, has suggested that the child is drinking in the world, better to construct his neural 'models' of the environment. And it is plain that a stunted organism is produced by depriving an infant of the rich diet of impressions on which his curiosity normally feeds with such extravagance. Animals raised in homogenized environments show crippling deficits in their later ability to learn and to transfer what they have learned. Children 'kept in the attic' by misguided or psychotic parents show the same striking backwardness. Indeed, even the children who have suffered the dull, aseptic environment of backward foundling homes often show a decline in intelligence that can be compensated only by vigorous measures of enrichment. So surely, then, an important early function is served by the child's omnivorous capacity for new impressions. He is sorting the world, storing those things that have some recurrent regularity and require 'knowing', discriminating them from the parade of random impressions. [See Bruner, 1959, on the functions of early curiosity.]

But if attention is to be sustained, directed to some task and held there in spite of temptations that come along, then obviously constraints must be established. The voluntary deployment of curiosity, so slowly and painfully mastered, seems to be supported in part by the young child's new-found capacity to 'instruct himself', literally to talk to himself through a sustained sequence. And in part the steadying force seems to be the momentum of concrete overt acts that have a way of sustaining the attention required for their completion by shutting off irrelevant impressions. In time, and with the development of habitual activities, and of language, there emerges more self-directed attention, sometimes called derived primary attention. The child is held steady not so much by vividness as by the habitual round of activity that now demands his attention. Little enough is known about how to help a child become master of his own attention, to sustain it over a long, connected sequence. But while young children are notoriously wandering in their attention, they can be kept in a state of rapt and prolonged attentiveness by being told compelling stories. There may be something to be learned from this observation. What makes the internal sequence of a story even more compelling than the distractions that lie outside it? Are there comparable properties inherent in other activities? Can these be used to train a child to sustain his curiosity beyond the moment's vividness?

Observe a child or group of children building a pile of blocks as high as they can get them. Their attention will be sustained to the flashing point until they reach the climax when the pile comes crashing down. They will return to build still higher. The drama of the task is only its minor virtue. More important is the energizing lure of uncertainty made personal by one's own effort to control it. It is almost the antithesis of the passive attraction of shininess and the vivid. To channel curiosity into more powerful intellectual pursuits requires precisely that there be this transition from the passive, receptive, episodic form of curiosity to the sustained and active form. There are games not only with objects, but with ideas and questions–like Twenty Questions–that provide such a disciplining of the channeling of curiosity. In so far as one may count on this important human motive–and it seems among the most reliable of the motives–then it seems obvious that our artificial education can in fact be made less artificial from a motivational standpoint by relating it initially to the more surfacy forms of curiosity and attention, and then cultivating curiosity to more subtle and active expression. I think it is fair to say that most of the success in contemporary curriculum building has been achieved by this route. When success comes, it takes the form of recognition that beyond the few things we know there is a domain of inference: that putting together the two and two that we have yields astonishing results. But this raises the issue of competence, to which we must turn next.

For curiosity is only one of the intrinsic motives for learning. The drive to achieve competence is another. White (1959) puts the issue well:

> According to Webster, competence means fitness or ability, and the suggested synonyms include capability, capacity, efficiency, proficiency, and skill. It is therefore a suitable word to describe such things as

> grasping and exploring, crawling and walking, attention and perception, all of which promote an effective–a competent–interaction with the environment. It is true, of course, that maturation plays a part in all these developments, but this part is heavily overshadowed by learning in all the more complex accomplishments like speech or skilled manipulation. I shall argue that it is necessary to make competence a motivational concept; there is *competence motivation* as well as competence in its more familiar sense of achieved capacity. The behavior that leads to the building up of effective grasping, handling, and letting go of objects, to take one example, is not random behavior that is produced by an overflow of energy. It is directed, selective, and persistent, and it continues not because it serves primary drives, which indeed it cannot serve until it is almost perfect, but because it satisfies an intrinsic need to deal with the environment.

Observations of young children and of the young of other species suggest that a good deal of their play must be understood as practice in coping with the environment. Primatologists describe, for example, how young female baboons cradle infant baboons in their arms long before they produce their own offspring. In fact, baboon play can be seen almost entirely as the practice of interpersonal skills. Unlike human children, baboons never play with objects, and this, the anthropologists believe, is connected with their inability to use tools when they grow up. And there is evidence that early language mastery, too, depends on such early preparation. One linguist (Weir, 1962) recently has shown how a two-year-old goes on exploring the limits of language use even after the lights are out, parents removed, communication stopped, and sleep imminent.

The child's metalinguistic play is hard to interpret as anything other than pleasure in practicing and developing a new skill. Although competence may not 'naturally' be directed toward school learning, it is certainly possible that the great access of energy that children experience when they 'get into a subject they like' is made of the same stuff.

We get interested in what we get good at. In general, it is difficult to sustain interest in an activity unless one achieves some degree of competence. Athletics is the activity par excellence where the young need no prodding to gain pleasure from an increase in skill, save where prematurely adult standards are imposed on little leagues formed too soon to ape the big ones. A custom introduced some years ago at the Gordonstoun School in Scotland has become legendary. In addition to conventionally competitive track and field events within the school, there was established a novel competition in which boys pitted themselves against their own best prior record in the events. Several American schools have picked up the idea and, while there has been no 'proper evaluation', it is said that the system creates great excitement and enormous effort on the part of the boys.

To achieve the sense of accomplishment requires a task that has some beginning and some terminus. Perhaps an experiment can serve again as a parable. There is a well-known phenomenon known to psychologists by the forbidding name of the Zeigarnik Effect. In brief, tasks that are interrupted are much more likely to be returned to and completed, and much more likely to be remembered, than comparable tasks that one has completed without interruption. But that puts the matter superficially, for it leaves out of account one factor that is crucial. The effect holds only if the tasks that the subject has been set are ones that have a structure–a beginning, a plan, and a terminus. If the tasks are 'silly' in the sense of being meaningless, arbitrary, and without visible means for checking progress, the drive to completion is not stimulated by interruption.

It seems likely that the desire to achieve competence follows the same rule. Unless there is some meaningful unity in what we are doing and some way of telling how we are doing, we are not very likely to strive to excel ourselves. Yet surely this too is only a small part of the story, for everybody does not want to be competent in the same activities, and some competencies might even be a source of embarrassment to their possessors. Boys do not thrill to the challenge of sewing a fine seam (again, in our culture), nor girls to becoming competent street fighters. There are competencies that are appropriate and activating for different ages, the two sexes, different social classes. But there are some things about competence motives that transcend these particulars. One is that an activity (given that it is 'approved'), must have some meaningful structure to it if it requires skill that is a little bit beyond that now possessed by the person–that it be learned by the exercise of effort. It is probably the combination of the two that is critical.

Experienced teachers who work with the newer curricula in science and mathematics report that they are surprised at the eagerness of students to push ahead to next steps in the course. Several of the teachers have suggested that the eagerness comes from increased confidence in one's ability to understand the material. Some of the students were having their first experience of understanding a topic in some depth, of going somewhere in a subject. It is this that is at the heart of competence motives, and surely our schools have not begun to tap this enormous reservoir of zest.

While we do not know the limits within which competence drives can be shaped and channeled by external reward, it seems quite likely that they are strongly open to external influence. But channelization aside, how can education keep alive and nourish a drive to competence–whether expressed in farming, football, or mathematics? What sustains a sense of pleasure and achievement in mastering things for their own sake–what Thorstein Veblen referred to as an instinct for workmanship? Do competence motives strengthen mainly on their exercise, in whatever context they may be exercised, or do they depend also upon being linked to drives for status, wealth, security, or fame?

There are, to begin with, striking differences among cultures and between strata within any particular society with respect to the encouragement given to competence drives. McClelland (1961), for example, in writing about the 'achieving society', comments upon the fact that in certain times and places one finds a flowering of achievement motivation strongly supported by the society and its institutions and myths alike. Emphasis upon individual responsibility and initiative, upon independence in decision and action, upon perfectibility of the self–all of these things serve to perpetuate more basic competency motives past childhood.

But cultures vary in their evaluation of *intellectual* mastery as a vehicle for the expression of competence. Bales (1944), for example, in comparing Irish and Jewish immigrant groups in Boston, remarks that the Jewish, much more than the Irish, treat school success and intellectuality as virtues in their own right as well as ways of upward mobility. The reasons can be found in history. Zborowski and Herzog (1952), in their book on eastern European Jewish communities, suggest that the barrier erected against Jews' entering other professions may have helped foster the cultivation of intellectual excellence as a prized expression of competence.

A culture does not 'manage' these matters consciously by the applications of rewards and reproofs alone. The son of the rabbi in the eastern European *stetl* was not punished if he wished to become a merchant rather than a Talmudic scholar, and, indeed, if he chose to become the latter he typically went through long, extrinsically unrewarding, and arduous training to do so. More subtle forces are at work, all of them fairly familiar but too often overlooked in discussing education. One of them is 'approval'. The professional man is more 'respected' than the manual worker. But that scarcely exhausts the matter. Respected by whom? Contemporary sociologists speak of the approval of one's 'reference group'–those to whom one looks for guides to action, for the definition of the possible, for ultimate approbation. But what leads *this* individual to look to *that* particular reference group?

What appears to be operative is a process we cavalierly call identification. The fact of identification is more easily described than explained. It refers to the strong human tendency to model one's 'self' and one's aspirations upon some other person. When we feel we have succeeded in 'being like' an identification figure, we derive pleasure from the achievement and, conversely, we suffer when we have 'let him down'. In so far as the identification figure is also 'a certain kind of person'–belongs to some group or category–we extend our loyalties from an individual to a reference group. In effect, then, identification relates one not only to individuals, but to one's society as well.

While this account is oversimplified, it serves to underline one important feature of identification as a process–its self-sustaining nature. For what it accomplishes is to pass over to the learner the control of punishment and reward. In so far as we now carry our standards with us, we achieve a certain independence from the immediate rewards and punishments meted out by others.

It has been remarked by psychologists that identification figures are most often those who control the scarce psychological resources that we most desire–love, approval, sustenance. Let me skip this issue for a moment and return to it later.

The term *identification* is usually reserved for those strong attachments where there is a considerable amount of emotional investment. But there are 'milder' forms of identification that are also important during the years of childhood and after. Perhaps we should call those who serve in these milder relationships 'competence models'. They are the 'on the job' heroes, the reliable ones with whom we can interact in some way. Indeed, they control a rare resource, some desired competence, but what is important is that the resource is attainable by interaction. The 'on the job' model is nowhere better illustrated than in the manner in which the child learns language from a parent. The tryout-correction-revision process continues until the child comes to learn the rules whereby sentences are generated and transformed appropriately. Finally he develops a set of productive habits that enable him to be his own sentence maker and his own corrector. He 'learns the rules of the language'. The parent is the model who, by interaction, teaches the skill of language.

In the process of teaching a skill the parent or teacher passes on much more. The teacher imparts attitudes toward a subject and, indeed, attitudes toward learning itself. What results may be quite inadvertent. Often, in our schools, for example, this first lesson is that learning has to do with remembering things when asked, with maintaining a certain undefined tidiness in what one does, with following a train of thought that comes from outside rather than from within and with honoring right answers. Observant anthropologists have suggested that the basic values of the early grades are a stylized version of the feminine role in the society, cautious rather than daring, governed by a ladylike politeness.

One recent study by Sears (1966) underlines the point. It suggests that girls in the early grades, who learn to control their fidgeting earlier and better than boys, are rewarded for excelling in their 'feminine' values. The reward can be almost too successful, so that in later years it is difficult to move girls beyond the orderly virtues they learned in their first school encounters. The boys, more fidgety in the first grade, get no such reward and as a consequence may be freer in their approach to learning in later grades. Far more would have to be known about the other conditions present in the lives of these children to draw a firm conclusion from the findings, but it is nonetheless suggestive. There are surely many ways to expand the range of competence models available to children. One is the use of a challenging master teacher, particularly in the early grades. And there is film or closed-circuit television, opening up enormously the range of teachers to whom the student can be exposed. Filmed teaching has, to be sure, marked limits, for the student cannot interact with an image. But a kind of pseudo interaction can be attained by including in the television lesson a group of students who are being taught right on the screen, with whom the student can take common cause. Team teaching provides still another approach to the exemplification of a range of competences, particularly if one of the teachers is charged specially with the role of gadfly. None of the above is yet a tried practice, but pedagogy, like economics and engineering, often must try techniques to find not only whether they work, but how they may be made to work.

I would like to suggest that what the teacher must be, to be an effective competence model, is a day-to-day working model with whom to interact. It is not so much that the teacher provides a model to *imitate*. Rather, it is that the teacher can become a part of the student's internal dialogue—somebody whose respect he wants, someone whose standards he wishes to make his own. It is like becoming a speaker of a language one shares with somebody. The language of that interaction becomes a part of oneself, and the standards of style and clarity that one adopts for that interaction become a part of one's own standards.

Finally, a word about one last intrinsic motive that bears closely upon the will to learn. Perhaps it should be called reciprocity. For it involves a deep human need to respond to others and to operate jointly with them toward an objective. One of the important insights of modern zoology is the importance of this intraspecies reciprocity for the survival of individual members of the species. The psychologist Barker (1963) has commented that the best way he has found to predict the behavior of the children whom he has been studying in great detail in the midst of their everyday activities is to know their situations. A child in a baseball game behaves baseball; in the drugstore the same child behaves drugstore. Situations have a demand value that appears to have very little to do with the motives that are operative. Surely it is not simply a 'motive to conform'; this is too great an abstraction. The man who is regulating his pressure on the back of a car, along with three or four others, trying to 'rock it out', is not so much conforming as 'fitting his efforts into an enterprise'. It is about as primitive an aspect of human behavior as we know.

Like the other activities we have been discussing, its exercise seems to be its sole reward. Probably it is the basis of human society, this response through reciprocity to other members of one's species. Where joint action is needed, where reciprocity is required for the group to attain an objective, then there seem to be processes that carry the individual along into learning, sweep him into a competence that is required in the setting of the group. We know precious little about this primitive motive to reciprocate, but what we do know is that it can furnish a driving force to learn as well. Human beings (and other species as well) fall into a pattern that is required by the goals and activities of the social group in which they find themselves. 'Imitation' is not the word for it, since it is usually not plain in most cases what is to be imitated. A much more interesting way of looking at what is involved is provided by the phenomenon of a young child learning to use the pronouns 'I' and 'you' correctly. The parent says to the child, 'You go to bed now'. The child says, 'No, you no go to bed'. We are amused. 'Not *me* but *you*', we say. In time, and after a surprisingly brief period of confusion, the child learns that 'you' refers to himself when another uses it, and to another person when he uses it–and the reverse with 'I'. It is a prime example of reciprocal learning. It is by much the same process that children learn the beautifully complicated games they play (adult and child games alike), that they learn their role in the family and in school, and finally that they come to take their role in the greater society.

The corpus of learning, using the word now as synonymous with knowledge, is reciprocal. A culture in its very nature is a set of values, skills, and ways of life that no one member of the society masters. Knowledge in this sense is like a rope, each strand of which extends no more than a few inches along its length, all being intertwined to give a solidity to the whole. The conduct of our educational system has been curiously blind to this interdependent nature of knowledge. We have 'teachers' and 'pupils', 'experts' and 'laymen'. But the community of learning is somehow overlooked.

What can most certainly be encouraged–and what is now being developed in the better high schools–is something approximating the give and take of a seminar in which discussion is the vehicle of instruction. This is reciprocity. But it requires recognition of one critically important matter: you cannot have both reciprocity and the demand that everybody learn the same thing or be 'completely' well rounded in the same way all the time. If reciprocally operative groups are to give support to learning by stimulating each person to join his efforts to a group, then we shall need tolerance for the specialized roles that develop–the critic, the innovator, the second helper, the cautionary. For it is from the cultivation of these interlocking roles that the participants get the sense of operating reciprocally in a group. Never mind that this pupil for this term in this seminar has a rather specialized task to perform. It will change. Meanwhile, if he can see how he contributes to the effectiveness of the group's operations on history or geometry or whatnot, he is likely to be the more activated. And surely one of the roles that will emerge is that of auxiliary teacher–let it, encourage it. It can only help in relieving the tedium of a classroom with one expert up here and the rest down there.

At the risk of being repetitious, let me restate the argument. It is this. The will to learn is an intrinsic motive, one that finds both its source and its reward in its own exercise. The will to learn becomes a 'problem' only under specialized circumstances like those of a school, where a curriculum is set, students confined, and a path fixed. The problem exists not so much in learning itself, but in the fact that what the school imposes often fails to enlist the natural energies that sustain spontaneous learning–curiosity, a desire for competence, aspiration to emulate a model, and a deep-sensed commitment to the web of social reciprocity. Our concern has been with how these energies may be cultivated in support of school learning. If we know little firmly, at least we are not without reasonable hypotheses about how to proceed. The practice of education does, at least, produce interesting hypotheses. After all, the Great Age of Discovery was made possible by men whose hypotheses were formed before they had developed a decent technique for measuring longitude.

You will have noted by now a considerable de-emphasis of 'extrinsic' rewards and punishments as factors in school learning. There has been in these pages a rather intentional neglect of the so-called Law of Effect, which holds that a reaction is more likely to be repeated if it has previously been followed by a 'satisfying state of affairs'. I am not unmindful of the notion of reinforcement. It is doubtful, only, that 'satisfying states of affairs' are *reliably* to be found outside learning itself–in kind or harsh words from the teacher, in grades and gold stars, in the absurdly abstract assurance to the high school student that his lifetime earnings will be better by 80 per cent if he graduates. External reinforcement may indeed get a particular act going and may even lead to its repetition, but it does not nourish, reliably, the long course of learning by which man slowly builds in his own way a serviceable model of what the world is and what it can be.

References

BAKER, R. (1963) 'On the Nature of the Environment', *J. social Issues, 19,* 17–38.

BALES, R. F. (1944) *The 'Fixation Factor' in Alcohol Addiction: A Hypothesis Derived from a Comparative Study of Irish and Jewish Social Norms,* unpublished Ph.D. thesis, Harvard University.

BRUNER, J. S. (1959) 'The cognitive consequences of early sensory deprivation', *Psychosomatic Medicine, 21,* 2, 89–95.

McCLELLAND, D. C. (1961) *The Achieving Society.* Princeton, New Jersey: Van Nostrand.

SEARS, P. (1966) 'Attitudinal and affective factors affecting children's approaches to problem solving', in BRUNER, J. S. (ed.) *Learning about Learning: A Conference Report.* Washington, D.C.: US Office of Education.

WEIR, R. H. (1962) *Language in the Crib.* The Hague: Mouton.

WHITE, R. W. (1959) 'Motivation Reconsidered: The Concept of Competence', *Psychol. Rev., 66,* 297–333.

ZBOROWSKI, M. and HERZOG, E. (1952) *Life is with People: The Jewish Little-Town of Eastern Europe.* New York: International Universities Press.

12 Social class and achievement motivation

D. F. Swift

The basic facts of social class performance in school are so well known as hardly to need repeating. As all teachers know, the children who do the best work, are easiest to control and stimulate, make the best prefects, stay at school longest, take part in extra-curricular activities, finish school with the best qualifications and references and get into the best jobs, tend to come from the middle class. The relationship exists in the junior school (Douglas, 1964), manifests itself in the 11+ (Floud *et al*., 1956; Fraser, 1959; Swift, 1965) and seriously affects secondary education (*Early Leaving*, Ministry of Education, 1954).

Swift (1965) found that the children of middle-class parents had six times as good a chance of selection at 11+ as working-class children. In this study it became clear that the lack of balance in these 'class chances' was mainly due to the fact that the sons of middle-class parents were, to that degree, better able to score on intelligence and attainment tests. Many psychologists explain this by suggesting that there is a basic intellectual superiority in the middle class which is maintained by selective mating and a tendency to inherit innate intelligence. This explanation begs as many questions as it answers, but their discussion is irrelevant to the material with which this paper will deal, since all parties would agree that there is room for some environmental distortion of the basic genetic relationships.

A recent study by a psychologist found that all the factors which are usually thought to be bad for the intellectual development of the child (except a working mother) were in fact related to poor performance in the school. Fraser (1959) found that parental education and reading habits, income, occupation and living space all related significantly with IQ and school performance. In the motivational sphere, parental attitudes towards the education and future employment of the child, together with the degree of encouragement which they felt they offered the child towards school work, were all significantly related with IQ and performance. Finally, abnormality in the family was found to be an important correlate of poor performance.

Of the eleven 'aspects' of the environment which Fraser considered, the nine mentioned above can be fairly called important elements of the social class situation. Of the remaining two, one (mother at work) was not significant, whilst the second (abnormal background) is to some extent a social class definition.

Since these 'aspects' are also important elements of the social class

---

Source: *Educational Research*, 1966, *8*, 83–95.

system a question immediately arises as to whether they would maintain their significance if the population were separated into groups representing different culture-patterns. Under these circumstances, would the 'aspects' retain their significance for performance, or does it derive from the fact that they are simply dimensions of a certain social class sub-culture? It is unfortunate that social class itself has not received the scientific 'mapping out' necessary to its being treated as a variable which can be strictly controlled, and it is possible to suggest that it never will, but this does not make it permissible to ignore it.

The sociological stream of thought which has provided the most valuable theory concerning the relationship between socio-cultural learning and the development of intellectual and motivational skills is the 'clash of values' approach which began with Allison Davis's (1948) inaugural lecture on social class influences upon learning. (See also Davis, 1949 and 1960.)

It was Davis's contention that there are barriers to understanding *within* societies which are every bit as serious as those which exist *between* societies; barriers which are all the more serious for the fact that the participating members do not realize their existence. Consequently, Davis enjoined educationists to study the social class system to discover those mental problems, solutions and cultural definitions of what is pleasant and desirable which are emphasized within the different social classes. Similarly, Toby (1957) analysed the problems involved in the maladjustment which occurs when the lower-class child brings the attitudes towards education of his culture into the school.

Child-rearing techniques, motivation and school achievement

The Davis critique of objective and subjective intelligence testing in the educational system stimulated research studies which can be grossly separated according to their concern for either child-rearing practices or parental value-systems. These two aspects of the learning process are clearly interdependent and to concentrate in research upon one aspect without the other is to run the risk of discovering irrelevant relationships since important factors in the other perspective are left uncontrolled. This becomes more important the 'nearer' the research approaches the individual. That is, the more detailed the definition of influential factors the greater the danger that its context is ignored. Child-rearing research has suffered greatly from this weakness and despite valiant efforts to take a synoptic view of the many non-comparable and occasionally contradictory findings (Bronfenbrenner, 1958; Clausen and Williams, 1963), headway has been slow. Research has concentrated upon a few of the more obvious mechanisms by which parents attempt to socialize children, and in doing this it has often simplified to a caricature the complex relationship between the environment and the developing personality.

Perhaps the grossest simplification undertaken by these studies is to assume that parental perceptions of what they do to a child are adequate representations of what *is* done to a child by the parent. Human interaction is far more subtle than this. A child is able to 'tune into' the implicit culture of his family life without too much interference from the manifest attitudes of parents towards what he should do in a particular situation. The folk-lore adage, 'do as I say, not as I do', comes nearer to an understanding of this complex relationship than many of the researches.

As an example of the contradictory conclusions which these studies have produced, we can compare Drews and Teaman (1957) with Fraser (1959). Drews discovered the attitudes of mothers of high and low academic achievers of both gifted and average intelligence on scales of permissiveness, protectiveness and domination. Her conclusions were that mothers of high achievers were more authoritarian and restrictive than mothers of low achievers. Parents of high achievers of gifted intelligence also seemed to have more punitive attitudes with respect to child-rearing. In contradiction, Fraser's research implies that the friendliness and spontaneity of a lenient 'democratic' atmosphere provided the 'best' (in terms of school motivation and achievement) family environment. Similarly, Elder (1965) investigated the association between perceived parent-adolescent relations and the likelihood of reaching secondary school in the United States, Great Britain, Western Germany, Italy and Mexico. He found a negative association between parental dominance and access to secondary education in all five countries where the structure of educational opportunity made it feasible.

Despite the contradictions in research findings, Bronfenbrenner, by dint of imaginative understanding and a not altogether persuasive hypothesis[1] perceived a change taking place during the twenty-five year period of research which he reviewed. For Bronfenbrenner, middle-class parents in America have changed from a 'restrictive' to a 'permissive' approach. Fortunately, he avoids the superficially reasonable conclusion that since there has been no change in middle-class ability to out-perform and out-drive the working class, child-rearing methods are of no importance. As Kohn (1963) points out, the researchers summarized by Bronfenbrenner were only as good as the questions they used and hence their implied conceptual models. The changes which took place represent faddish alterations in superficial techniques which are not symptomatic of changes in the basic relationships between parents and children. 'Parents have changed techniques in the service of much the same values' (Kohn 1963, p. 473).

A further point of view has been brought to bear on the question by Veroff (1965), who suggests a theoretical background to the problem in motivational research of distinguishing between situational and personality factors in any situation: Veroff's major hypothesis is:

---

[1] That middle-class parents are more responsive to mass media child-rearing instruction.

> that there is a critical stage for learning the affective associations attendant to the development of motives and motivational expectancies. If the environment supports learning these associations during the critical period, then strong motivational dispositions will develop; if the environment supports a time earlier than the critical time, anxiety connected with these motivations will develop and if the environment supports learning at a date later than the critical time, very specific motives and expectancies will ensue (p. 17).

This emphasizes the specific contribution of the environment. Furthermore, it implies that, in research, we should treat the environment as a whole which has a specific meaning for the developing personality and intellect deriving from the specific configurations of the whole matrix of factors. This is an important point which may help to explain the inconsistencies of research findings. The usual psychological approach has been to isolate particular factors (specific techniques in punishment for example) and relate these to motivation or achievement. It was not understood that this specific research act carried an implicit assumption that the same techniques carried out in this way meant the same thing to the child regardless of the social context. It was thought of as a single 'stimulus' similar to the stimulus which is produced in a laboratory by manipulation of an independent variable. Unfortunately (or fortunately), family life is not a laboratory and this assumption cannot be made at present.[2] The 'situation' must be seen as a matrix of interrelating factors which is destroyed when a single crude factor is abstracted since the factor derives its meaning from its context. For example, 'but you *must* go to school', does not mean the same in a context of love and mutual respect as it does in a context in which the child fully understands the implicit relief of the parent in being rid of him. Yet the degree of firmness might be exactly the same.

Because of this problem in dealing with the environment,[3] it is to be expected that those studies which looked for a generalized picture of the family environment ('loving', 'supportive', etc.) and those which employed social class analysis, came closest to general agreement.

Before dealing with social class analysis we might mention a fairly elementary relationship which does appear to hold in all situations. Parental encouragement does have an important bearing upon school motivation and performance. Mannino (1962) reported that, with socio-economic status controlled, 'staying on' at school after the minimum leaving age related positively to parental encouragement. Wall and Miller (1962) concluded that 'children, particularly boys, do very much better at school if their parents are interested in their progress', and that 'children whose parents are rated as "very interested" improve their test-scores between eight and eleven but those whose parents are "uninterested" show a deterioration in score'. Douglas (1964, p. 61) also suggested that children tend to work well when their parents take an interest in their school progress and to work badly when they are uninterested.

---

[2] This is not to say that the logic of research is wrong. The ideal aim is still to discover in the environment all the most important stimuli and to refine the techniques by which they can be measured. It only means that our actual abilities to do this in family life research covering a long period of time are incapable of handling the environment in this way at the present time.

[3] As a sociologist the writer must insist upon including the individual's perception of the social reality as part of 'the environment' since psychologists often do their research studies serious damage by assuming that 'the environment' constituted only the material and geographical aspects of the environment.

Social class values and achievement motivation

Many studies have shown a positive association between the level of educational and occupational aspirations on the one hand and various measures of social status on the other. Usually it is assumed that this association is a 'real' one which results from the influence of the particular constellations of occupational, educational and action values which are implicit in the culture of the middle class. However, the equally clear relationship between school ability and middle-class status is not usually accepted at face value. At least part of this association is thought to be due to an intervening variable called 'intelligence'. Sewell *et al.* (1957) considered the possibility that the aspiration-social class relationship was also due to the intervention of this factor, intelligence. Tests on a random sample of high school seniors covering the state of Wisconsin confirmed that the relationship between aspiration and social class was not simply due to the common relationship of these variables to intelligence. Social class (whatever that is) makes an independent contribution. This is important because the conclusion was arrived at by controlling for 'intelligence', which itself can be shown to be derived, to some extent, from socio-cultural learning experiences. Thus, the importance of social status showed through even though it was weakened by controlling intelligence.

With Sewell as justification and Davis as inspiration we can now consider the value-systems of the different social classes to be learning environments which participate in the development of motivation and achievement levels. Davis had urged the use of class analysis in investigating the environment for factors in intellectual and personality development because it appears to provide the best means by which the cultural experience of the child may be described. It represents the best shorthand method for summarizing the individual's cultural beliefs, his definitions of life problems and possible solutions, the meaning of his language, and his culturally learned conceptions of the teacher, himself and school. As Kohn (1963, p. 471) puts it: 'It is so useful because it captures the reality that the intricate interplay of all these variables creates different basic conditions of life at different levels of the social order.'

Early Leaving was taking a similar, if much cruder, line when, in partial explanation of the fact that working-class children leave grammar school before their middle-class peers, it has the following to say:

> One of the main influences must be sought in the outlook and assumptions of parents in various walks of life. Consider first the outlook of parents in professional occupations. Most of them themselves received a grammar school or similar education, and others have made their way into a position in society in which they find such a background taken for granted. They are all engaged on work for which a fairly high level of education is an obvious advantage and many follow professions to which a specified educational standard is a condition of entry. In the circumstances it is not surprising that they assume that their children will not leave school at 15 but will stay as much longer as their ability justifies. This assumption is not due to any conscious sense of the value of education; it may be a mere social convention which has never been questioned. But it is in any case a powerful influence on parents towards keeping their children at school, and on children towards staying (p. 36).

Before *Early Leaving* was published, Kahl (1953) documented a similar phenomenon in the American educational system. During the course of a nationwide mobility study it became plain that 'IQ and family status were useful predictors of the educational and occupational ambitions of high school boys'. However, within the lower-middle range it was found that these relationships were weaker than at the higher and lower reaches of the scale. Kahl, therefore, chose a pair of matched samples of boys from this status group, one representing boys who wished to go to college and another representing boys who did not wish to go. Interviews revealed that, 'although there was a common way of life which identified the common-man class, some members were content with that way of life while others were not'. Discontented parents, seeing that the educational system was a way out for their sons, tended to train them from the earliest years of grammar school to take school seriously. 'Only sons who internalized such values were sufficiently motivated to overcome the obstacles which faced the common-man boys in school; only they saw a good reason for good school performance and college aspirations' (Kahl, 1953, p. 203).

An important finding of the study was that boys, to an extent not previously imagined, took up and held as their own, their parents' perceptions of possible opportunities, the desirability and possibility of status mobility, methods of obtaining mobility, and the goals which they felt appropriate to boys in their position at school. Neither the school nor mass media seemed to have much influence in moulding these ideas. This finding, which has found support in other research, is a very firm invitation to researchers to concentrate upon family environment, and since that time a great deal of research has been carried out on parental values and their effect upon motivation in school.

A further extension of this work within the Harvard Mobility Survey was undertaken by Elizabeth Cohen (1959) who studied fifty matched pairs of boys in the 'common-man' group. Each pair consisted of a 'mobile' and a 'non-mobile' boy matched for IQ, ethnicity, town of origin and parental membership of the common-man group. Attitudes of parents and sons were obtained by Guttman scale testing, and the basic aim was to chart the channels of parental influence upon the motivation of their sons. Cohen's theoretical model included three principal channels of influence; through socio-economic characteristics, through generalized attitudes towards college, and through concrete encouragement. In these ways the parents were expected to influence the decisions of their sons as to whether they wished to go to college or not.

In analysing the influence of socio-economic status, Cohen suggested that a significant difference in motivation is found among children of working-class parents who have close 'on the job' associations with people of higher status. As regards 'pressure' it was found that parents of mobile sons seemed deliberately to encourage them to aim for middle-class occupations and educational qualifications but did not push for higher school performance as such.

A similar sort of study was reported by Reissman (1953), aimed at determining the relationship between class and aspirations. Apart from age, he found that the two important aspects of the individual's value-system were his reference group and certain 'life-orientations'. The reference group is obviously important to the degree of satisfaction the individual has with past mobility and future aims. Thus a low level of aspiration can be based on a feeling of achievement in relation to the normal achievement of his reference group. Orientations other than success and mobility can be dominant for some individuals and thereby lead them to exert their energies in a different direction, or prevent them from aiming for mobility.

In England, Himmelweit (1955) enquired into the relationship between socio-economic background and behaviour on the one hand, and the formation of attitudes and values on the other. It is interesting to note here the use to which a social psychologist has put the Davis hypothesis. For Dr Himmelweit and many American social psychologists, this hypothesis has suggested that 'a kind of adaptive socialized anxiety is generated more rapidly as a result of middle-class methods of upbringing'. The difficulty experienced by working-class children to adapt to the middle-class school culture was also considered. Information was collected on a wide variety of topics with the aim of ascertaining the degree to which Davis's hypothesis could be confirmed in England on representative samples of grammar school and secondary modern school children from middle-middle to lower working class.

The results confirmed some of Davis's and Warner's hypotheses.

Middle-class children were more concerned with how well they do at school, as are their parents. Parents also supervise the leisure of their children more. However–

> despite so many confirmatory results as to the greater pressure to which the middle-class child is subjected, both externally by the parents and internally by his more rigid super-ego and more intra-punitive technique of handling frustrations, no evidence of greater overall anxiety or tension was obtained (Himmelweit, 1955, p. 34).

A further set of findings was concerned with the transmission of values. One of the hypotheses was that the upwardly mobile working-class boy would try to become like his middle-class co-pupil with the result that his answers might be different from those of the non-mobile members of his class. It was also expected that the two middle-class groups (successful and unsuccessful) would be similar. There were two main results. Firstly, the difference between the school groups of the middle and working classes were so much greater than those between social classes within the same school that differences were thought to reflect, in large measure, differences in educational and intellectual achievement. Secondly, the difference between the working-class groups was most marked. Parents of successful working-class boys had a much more middle-class set of values than even the middle class.

Hyman (1953) saw a system of beliefs and values peculiar to the lower class acting as an 'intervening variable mediating the relationship between low position and lack of upward mobility'. The components of this value-system were:

1. less emphasis on traditional high success goals;
2. increased awareness of lack of opportunity to achieve success;
3. less emphasis on the achievement of goals which in turn would be instrumental for success.

These all add up to a self-imposed barrier to mobility. Thus, apart from the structurally imposed barriers to mobility, they have a further set of hindrances in their own ideas about mobility.

Hyman analysed census and other data looking for evidence that motivation to advance to high places and to obtain the right sort of training for doing so was lower in the lower class. Considering first the value placed on formal education, he quoted four studies which supported the idea that the lower class always placed a lower value on education than those above them in the social scale. He found that there were crucial differences in values as regards motivation to advance in the economic structure. For example, when respondents were asked to choose their ideal occupation there was a strong tendency for upper-class people to choose occupations for 'congeniality to person' reasons. On the other hand, the choices of lower-class people indicated much more interest in the economic benefits of the occupation. Only a very few wealthy people chose skilled occupations as being preferable to others, whereas a major proportion of lower-class people chose this form of qualification. In these ways Hyman emphasized that there is reduced striving among the lower classes, an awareness of lack of opportunity, and different ideas as to how social mobility might come about.

The special relevance of values in the family was clarified by Rosen (1956) who expressed dissatisfaction with current psychological and sociological explanations of social mobility. Hitherto, says Rosen, the greater achievement of the middle class has been explained either by a process of social selection–the middle class being more intelligent, healthier and more attractive; or by the differential opportunities open to the two classes–the 'life chances' hypothesis. These explanations, he agreed, were relevant and consistent with each other, but they were not exhaustive:

> since neither explanation takes into account the possibility that there may be psychological and cultural factors which affect social mobility by influencing the individual's willingness to develop and exploit his talent, intelligence and opportunities (Rosen, 1956, p. 203).

Therefore, his thesis was that differential rates of mobility may also be explained as a function of differences in the motives and values of the social classes. Rosen saw that there were two components in the achievement orientations of the different classes. Firstly, a psychological factor–the personality characteristic of 'achievement motivation', which provides an internal impetus to excel. Secondly, there is also a cultural factor consisting of certain value orientations which define and implement achievement motivated behaviour. He suggested that values affect mobility in three ways:

1. they provide a definition of goals;
2. they focus the attention of the individual on achievement;
3. they prepare him to translate motive into action.

Thus, whether or not the individual tries for mobility will be decided by his values. Before the drive for achievement can be expressed in the right, culturally defined terms there must be awareness of, and willingness to undertake, the necessary steps for achievement. These will of necessity involve the ability to plan, to sacrifice, and to strive consistently. Here again, abilities like these will be partly dependent on the values of the individual.

Rosen goes further to make a crucial distinction between widespread values and implementary values. Frequently we can see no difference between classes in the holding of some popular widespread value like the desirability of upward mobility. But a difference is found in the implementary values which facilitate and encourage achievement. He summarized the situation in America in terms of basic or implementary values in this way:

> Middle-class children are more likely . . . to embrace the achievement value-system which states that given the willingness to work hard, plan and make proper sacrifices, an individual child should be able to manipulate his environment so as to ensure eventual success (Rosen, 1956, p. 211).

Progress, that is, is not necessarily a question of drive. At least as important is the way in which the individual fits in with the requirements of his present situation. This is obviously so with our education system. Successful mobility can come to those who simply adapt themselves well to each successive stage, the system itself carrying them up as on the crest of a wave. This meshing in with the values, perceptions and behaviours of a culture-pattern constitutes, with the achievement drive, the two basic factors which produce mobility within the framework laid down by the structure of possibilities.

Having established the abundance of theoretical justifications for an analysis of family value-patterns along social class lines, it now remains for us to deal with some of the more recent substantive findings of research studies which have been designed on these lines.

Douvan (1956) sets the stage with an experiment designed to test the hypothesis that 'achievement want' of middle-class children is more generalized than that of working-class children and hence is less likely to vary with changes in the reward offered for achievement. Her conclusion was that the achievement want of the middle-class child was more generalized as well as more dependable.

Douvan and Adelson (1958) also provide us with a useful thumbnail sketch of the upward- and downward-aspiring boy. In this survey, the authors classified boys as upward mobile, stable, or downward mobile, depending upon whether their occupational aspirations were higher than, similar to, or lower than their father's occupation. Interviews with the boys revealed the following categorizations. The upward-aspiring boy was described as being unusually lively and energetic. He expressed concern for the intrinsic interest of his future work and has a longer time-perspective when thinking of occupations. A set of pictures was used to discover the degree and form of internalized standards possessed by the two extreme groups. These pictures portrayed an adolescent boy facing conflicting pressures. For example, he might be torn between a promise to parents to be home at a particular hour and a wish to accept pressure from his friends to stay out late. In both groups two-thirds of the respondents decided that the boy would go home. Thus there was no difference in prognosticated *behaviour.* However, there was a difference in perceived motives for the action in that the upward-aspiring boys tended to explain his actions in terms of his promise to his parents, or because his parents had trusted him. The downward-aspirers were more likely to explain his return by fear of punishment. The implications of these findings for the authors was that the upward-oriented have internalized personal controls in contrast to the downward-oriented who showed a tendency to internalize external standards and rebel against them. The upward-aspirer was found to show greater independence from his family whereas the downward-aspirer was at once more dependent on, and rebellious towards, his family. Finally, the upward-aspirer showed high self-acceptance and social confidence in comparison with the downward-aspirer who was more ambivalent towards himself and more unsure in social situations.

These findings fit neatly with those of Kohn (1959) who showed clearly that child-rearing values of middle-class parents stress internalized standards for governing relationships with other people while the working class concentrated upon qualities of honesty and neatness. Apart from demonstrating the importance of social class analysis in this question (as do Kohn and Carroll, 1960) the findings support Duvall's description of working- and lower-middle-class parental values as 'traditional' in comparison with the middle-class 'developmental' emphasis. With the working-class parent it is the actual act that matters. The child must accept externally imposed rules. To the middle-class parent it is the motives of the child which give meaning to the act. Thus we can expect that where the middle-class family is successful in transmitting such a point of view to the child, he will develop a responsibility for his own values and actions which would support his acceptance of the parental views on educational and occupational aspirations. On the other hand, a working-class point of view on education (most social classes feel that education is A GOOD THING) will tend to be imposed upon the child together with other external rules. There is greater likelihood that the child will rebel against this as against other arbitrarily imposed rules. There is then a further factor that even where the working-class child internalizes the value it will mean something different to him. Horizons will be lower and understanding of the situation and its requirements less efficient.

Katz (1964) studied the meaning of success among Australian adolescents and found evidence of considerable variation according to socio-economic class of origin. These differences appeared to indicate 'a major difference in the frame of aspirational reference internalized by members of different social classes'. In the case of the middle-class adolescent, the success goal was prestige achievement attained through personal effort and worthiness. Adolescents from the unskilled working class, on the other hand, had little concern for social status. Success is defined in terms of 'posessions which are procured by personal exertion, but limited in possibility of attainment by factors over which the individual has no control'. Between these two extremes, adolescents of skilled worker families showed a great deal of variation, suggesting to Katz that some had internalized middle-class and others, unskilled-class, perceptions.

The importance of chance in the lower working-class perception of the environment is supported by research in Britain and most relevantly by the theories and research of Basil Bernstein (1962). On the basis of his analysis of linguistic forms, Bernstein suggests that middle-class life and language encourage the development of an 'instrumental' attitude to the world around which is distinct from the more intellectually passive perceptions of the working class.

We have seen that the values of working-class life tend to demand adoption of rules which are seen by the child to be external and arbitrary ('because I say so'). We also know that the horizons of social actuality implicit in the specific behavioural imperatives by which working-class parents seek to organize the behaviour of children are 'shorter' and 'narrower' than those of the middle-class families [though none the less 'accurate' for that (Liversidge, 1962)]. Since middle-class children have higher need achievement in relation to education and this clearly is important in actual performance,[4] we can assume that the imposition of arbitrary rules and shorter horizons are both depressive factors in the need for educational achievement. They clearly are when taken together and in certain cultural situations. However, they need not go together. In line with the earlier theoretical objections to abstraction of a single factor for treatment as a continuously operating variable, we have to consider the possibility that different situations will give different functions to any single factor.

The lower middle class provides a useful test of this possibility. Duvall (1946) had joined the lower middle class to the working class in categorizing their child-rearing ethos as 'traditional'. On the other hand, the work situation, upon which one might expect their achievement values to depend, would be different. They would tend to do work which demanded skills in the manipulation of ideas and persons as opposed to working-class activities with things. Promotion will, superficially at any rate, appear to depend upon the actual performance of the individual. Finally, the lower middle class is likely to be less subject to control than the working class.

With all these points in mind, and in the light of Argyle and Robinson's (1962) suggestion that introjection or projection of parental ambition is important in parent–child relationships, Swift (1964) related the 11+ performance of lower-middle-class children to the mobility pessimism of their fathers. In a ten per cent random sample of a local education authority division it was found that, within the middle class, the father's dissatisfaction with his job and its prospects related significantly to the likelihood of his child's success on the 11+. In addition it was found that mobility-pessimism was very dependent upon lower-middle-class membership. In the sample only one out of nine lower-middle-class fathers of successful children did not have 'high' mobility-pessimism.

---

[4] Douglas (1964, p. 60) found that according to teachers, 26 per cent of upper-middle-class children work 'very hard' as opposed to 7 per cent of upper and lower working-class: (p. 64) children who work 'hard' or 'very hard' get 12 per cent more grammar school places than they would be expected on their tests to get, while 'poor' or 'average' workers got 25 per cent fewer places than their tests would lead us to expect.

On the basis of these very clear but necessarily tentative findings, a second ideal-type of academically successful family was constructed.[5] The principal intention behind this was to emphasize the weakness of the position which tends to see 'the family environment' as a variable which ranges from 'good' to 'bad' in terms of its capacity to produce children who are adaptive to the needs of the educational system. This is a logically reasonable perspective to hold but it carries with it the danger of implicit assumptions about the linearity of relationships between its component elements and educational adaptiveness which are not necessarily valid. In this case, research has emphasized the importance of a supportive, 'democratic' family life which encourages the development of personal confidence and satisfaction with values which are accepted rather than imposed. The assumption which tends to follow from this is that the 'developmental' environment is best at the top and non-existent at the bottom of the social class hierarchy. In this way it can be seen to be a stimulus which relates in a linear fashion to need for educational achievement and hence educational success.

This second ideal-type of academically successful family was one in which mobility-pessimism was high, and discipline was 'traditional'. Commitment to education was 'high' but of a very different nature to the high commitment expressed in the ideal-type middle-class family where it is valued for its liberating qualities for the individual. Instead (in the light of the 'traditional' discipline ethos) it was seen as a necessary aspect of certificate-collecting which was, in turn, seen to be the major 'external' demand of the social environment. One way in which this family differs from the working-class family lies in the social horizons which the parents hold for their children. Because of similarities in work situation and content, and probably of early educational experience, the lower-middle-class parent will have horizons which are closer to those found in the remainder of the middle class. He is also likely to understand what is involved in school work and will not be burdened by the idea that the people who get into the higher classes represent a different sort of person from himself.[6]

---

[5] An alternative, that is, to the usual ideal-type of 'cultured' upper-middle-class family in which the child is stimulated in a generally tolerant fashion to develop his capabilities in many directions.

[6] For the reasons which Hyman (1953) made clear, this has a depressing effect upon aspiration and action.

This all adds up to a very different sort of environment to the stereotypical 'educationally successful' middle-class family. It is also one which cannot logically be placed midway between skilled working and middle class except on a single specific criterion like income. According to the hypothesis, the nearer a family approaches this ideal-type, the more likely is their child to be successful in the school. The whole analytical situation is then made vastly more complicated when we remember that the hypothesis has been framed in terms of two ideal-types, neither of which is likely to exist. Any given family will simultaneously go some way towards both ideal-types. However, social class analysis can help in 'controlling' for certain configurations of variables to the extent that work-situation, income and education tend to encourage their development.

Conclusion

An attempt has been made to present a perspective on the relationship between family environment and the need for educational achievement. This is a sociological perspective which employs the concept of values as the link between the social environment on the one hand and educational motivation on the other. It is suggested that research upon the development of educational potential has tended to oversimplify the subtle relationships between the developing personality and its environment.

However, research and theory which recognized the coherence of the cultural environment have suggested some general propositions. The accuracy of these propositions may not prove to be high, but if British research follows the lead given by American workers, away from descriptive and naive correlational studies towards socio-cultural analysis, there is real prospect of light.

References

ARGYLE, M. and ROBINSON, J. (1962) 'Two origins of achievement motivation', *Br. J. soc. clin. Psychol., 1,* part 2.

BENE, E. (1959) 'Some differences between middle class and working class grammar school boys in their attitudes towards education', *Br. J. Sociol., 10.*

BERNSTEIN, B. (1962) 'Social class linguistic codes and grammatical elements', *Language and Speech, 5,* 221–40.

BRONFENBRENNER, U. (1958) 'Socialization and social class through time and space', in E. E. MACCOBY, T. M. NEWCOMB, and R. L. HARTLEY (eds.) *Readings in Social Psychology.* New York: Henry Holt, pp. 400–25.

CAMPBELL, W. J. (1955) 'The influence of socio-cultural environment on the progress of children at the secondary school level', *Aust. J. Psychol.*, *7*, 140–6.

CENTERS, R. (1948) 'Motivational aspects of occupational stratification', *J. soc. Psychol.*, *28*, 187–214.

CLAUSEN, J. A. and WILLIAMS, J. (1963) 'Sociological correlates of child behaviour', in H. W. STEVENSON, J. KEGAN, and C. SPIKER (eds.) Child Psychology section in *Yearbook of National Society for Studies in Education*, *62*, 62–107.

COHEN, E. (1959) 'Parental factors in educational mobility', unpublished Ph.D. thesis, Harvard University.

COHEN, E. (1965) 'Parental factors in educational mobility', *Sociol. of Educ.*, *38*, no. 5, 401–25.

DAVIS, A. (1948) *Social Class Influences Upon Learning.* Harvard University Press.

DAVIS, A. (1949) 'American status systems and the socialization of the child', in C. KLUCKHOHN and H. A. MURRAY, *Personality in Nature, Society and Culture.* San Diego: Knopf.

DAVIS, A. (1960) *Psychology of the Child in the Middle Class.* University of Pittsburg Press.

DOUGLAS, J. W. B. (1964) *The Home and the School.* London: MacGibbon & Kee, pp. 60–5 and 169–62.

DOUVAN, E. (1956) 'Social status and success strivings', *J. abnorm. soc Psychol.*, *52*, 219–23.

DOUVAN, E. and ADELSON, J. (1958) 'The psychodynamics of social mobility in adolescent boys', *J. abnorm. soc. Psychol.*, *56*, 31–44.

DREWS, E. M. and TEAMAN, J. E. (1957) 'Parental attitudes and academic achievement', *J. clin. Psychol.*, *13*, 328–32.

DUVALL, E. M. (1946) 'Conceptions of parenthood', *Am. J. Sociol.*, *52*, 193–203.

ELDER, G. H. (1965) 'Family structure and educational attainment; a cross-national analysis', *Am. sociol. Rev.*, *30*, no. 1, 81–96.

FLOUD, J. E., HALSEY, A. H. and MARTIN, A. M. (1956) *Social Class and Educational Opportunities.* London: Heinemann.

FRASER, E. (1959) *Home Environment and the School.* London: University of London Press.

HIMMELWEIT, H. T. (1955) 'The Psychological aspects of social differentiation', *Inter. Soc. Science Bull.*, *7*.

HYMAN, H. H. (1953) 'The value systems of different classes', in R. BENDIX and S. LIPSET (eds.) *Class, Status and Power*, pp. 426–42.

KAHL, J. A. (1953) 'Educational and occupational aspirations of "common man" boys', *Harvard Educ. Rev.*, *23*, 186–203.

KATZ, F. M. (1964) 'The meaning of success: some differences in value systems of social classes', *J. soc. Psychol.*, *62*, 141–8.

KOHN, M. (1959) 'Social class and parental values', *Am. J. Sociol.*, *64*, 337–51.

KOHN, M. (1963) 'Social class and parent–child relationships: an interpretation', *Am. J. Sociol.*, *69*, 471–80.

KOHN, M. L. and CARROLL, E. E. (1960) 'Social class and the allocation of parental responsibilities', *Sociometry*, *23–24*, 372–92.

LIVERSIDGE, W. (1962) 'Life chances', *Sociol. Rev., 10,* no. 1, 17–34.

MANNINO, F. V. (1962) 'Family factors related to school persistence', *J. educ. Sociol., 35,* no. 5, 103–202.

MINISTRY OF EDUCATION (1954) *Early Leaving.* London: HMSO.

REISSMAN, L. (1953) 'Levels of aspiration and social class', *Am. sociol. Rev., 18,* 233–42.

ROSEN, B. C. (1956) 'The achievement syndrome: a psycho-cultural dimension of social stratification', *Am. sociol. Rev., 21,* 203–11.

ROSEN, B. C. and D'ANDRADE, R. (1959) 'The psycho-social origins of achievement motivation', *Sociometry, 22,* 185–218.

SEWELL, W. H., HALLER, A. O. and MURRAY, A. (1957) 'Social status and educational and occupational aspiration', *Am. sociol. Rev., 22,* 67–73.

SWIFT, D. F. (1964) 'Who passes the 11+?', *New Society,* 5 Mar., pp. 6–9.

SWIFT, D. F. (1965) 'Meritocratic and social class selection at age eleven', *Educ. Res., 8,* no. 1, 65–73.

TOBY, J. (1957) 'Orientation to education as a factor in school maladjustment of lower class children', *Social Forces, 35,* 259–66.

VEROFF, J. (1965) 'Theoretical background for studying the origins of human motivation dispositions', *Merrill-Palmer Quarterly of Behaviour and Development, 2,* no. 1.

WALL, W. D. and MILLER, K. M. (1962) 'Motivation and counter-motivation', in *Child and Education,* proceedings of XIV International Congress of Applied Psychology, vol. iii, 161–75.

13 The science of learning and the art of teaching

B. F. Skinner

Schoolroom teaching

Let us consider, for example, the teaching of arithmetic in the lower grades. The school is concerned with imparting to the child a large number of responses of a special sort. The responses are all verbal. They consist of speaking and writing certain words, figures, and signs which, to put it roughly, refer to numbers and to arithmetic operations. [Obviously this is not the 'new math', but a similar analysis might be made of any material suitable for the same grades.] The first task is to shape these responses–to get the child to pronounce and to write responses correctly–but the principal task is to bring this behavior under many sorts of stimulus control. This is what happens when the child learns to count, to recite tables, to count while ticking off the items in an assemblage of objects, to respond to spoken or written numbers by saying 'odd', 'even', or 'prime'. Over and above this elaborate repertoire of numerical behavior, most of which is often dismissed as the product of rote learning, the teaching of arithmetic looks forward to those complex serial arrangements of responses involved in original mathematical thinking. The child must acquire responses of transposing, clearing fractions, and so on, which modify the order or pattern of the original material so that the response called a solution is eventually made possible.

Now, how is this extremely complicated verbal repertoire set up? In the first place, what reinforcements are used? Fifty years ago the answer would have been clear. At that time educational control was still frankly aversive. The child read numbers, copied numbers, memorized tables, and performed operations upon numbers to escape the threat of the birch rod or cane. Some positive reinforcements were perhaps eventually derived from the increased efficiency of the child in the field of arithmetic and in rare cases some automatic reinforcement may have resulted from the sheer manipulation of the medium–from the solution of problems or the discovery of the intricacies of the number system. But for the immediate purposes of education the child acted to avoid or escape punishment. It was part of the reform movement known as progressive education to make the positive consequences more immediately effective, but any one who visits the lower grades of the average school today will observe that a change has been made,

---

Source: SKINNER, B. F. (1968) *The Technology of Teaching.* New York: Appleton-Century-Crofts, 14–21.

not from aversive to positive control, but from one form of aversive stimulation to another. The child at his desk, filling in his workbook, is behaving primarily to escape from the threat of a series of minor aversive events—the teacher's displeasure, the criticism or ridicule of his classmates, an ignominious showing in a competition, low marks, a trip to the office 'to be talked to' by the principal, or a word to the parent who may still resort to the birch rod. In this welter of aversive consequences, getting the right answer is in itself an insignificant event, any effect of which is lost amid the anxieties, the boredom, and the aggressions which are the inevitable by-products of aversive control.

Secondly, we have to ask how the contingencies of reinforcement are arranged. When is a numerical operation reinforced as 'right'? Eventually, of course, the pupil may be able to check his own answers and achieve some sort of automatic reinforcement, but in the early stages the reinforcement of being right is usually accorded by the teacher. The contingencies she provides are far from optimal. It can easily be demonstrated that, unless explicit mediating behavior has been set up, the lapse of only a few seconds, between response and reinforcement destroys most of the effect. In a typical classroom, nevertheless, long periods of time customarily elapse. The teacher may walk up and down the aisle, for example, while the class is working on a sheet of problems, pausing here and there to call an answer right or wrong. Many minutes intervene between the child's response and the teacher's reinforcement. In many cases—for example, when papers are taken home to be corrected—as much as twenty-four hours may intervene. It is surprising that this system has any effect whatsoever.

A third notable shortcoming is the lack of a skilful program which moves forward through a series of progressive approximations to the final complex behavior desired. A long series of contingencies is necessary to bring the pupil into the possession of mathematical behavior most efficiently. But the teacher is seldom able to reinforce at each step in such a series because she cannot deal with the pupil's responses one at a time. It is usually necessary to reinforce the behavior in blocks of responses—as in correcting a worksheet or page from a workbook. The responses within such a block must not be interrelated. The answer to one problem must not depend upon the answer to another. The number of stages through which one may progressively approach a complex pattern of behavior is therefore small, and the task so much the more difficult. Even the most modern workbook in beginning arithmetic is far from exemplifying an efficient *program* for shaping mathematical behavior.

Perhaps the most serious criticism of the current classroom is the relative infrequency of reinforcement. Since the pupil is usually dependent upon the teacher for being told that he is right, and since many pupils are usually dependent upon the same teacher, the total number of contingencies which may be arranged during, say, the first four years, is of the order of only a few thousand. But a very rough estimate suggests that efficient mathematical behavior at this level requires something of the order of 25,000 contingencies. We may suppose that even in the brighter student a given contingency must be arranged several times to place the behavior well in hand. The responses to be set up are not simply the various items in tables of addition, subtraction, multiplication, and division; we have also to consider the *alternative* forms in which each item may be stated. To the learning of such material we should add hundreds of responses such as those concerned with factoring, identifying primes, memorizing series, using short-cut techniques of calculation, and constructing and using geometric representations or number forms. Over and above all this, the whole mathematical repertoire must be brought under the control of concrete problems of considerable variety. Perhaps 50,000 contingencies is a more conservative estimate. In this frame of reference the daily assignment in arithmetic seems pitifully meagre.

The result of all this is, of course, well known. Even our best schools are under criticism for their inefficiency in the teaching of drill subjects such as arithmetic. The condition in the average school is a matter of widespread national concern. Modern children simply do not learn arithmetic quickly or well. Nor is the result simply incompetence. The very subjects in which modern techniques are weakest are those in which failure is most conspicuous, and in the wake of an ever-growing incompetence come the anxieties, uncertainties, and aggressions which in their turn present other problems to the school. Most pupils soon claim the asylum of not being 'ready' for arithmetic at a given level or, eventually, of not having a mathematical mind. Such explanations are readily seized upon by defensive teachers and parents. Few pupils ever reach the stage at which automatic reinforcements follow as the natural consequences of mathematical behavior. On the contrary, the figures and symbols of mathematics have become standard emotional stimuli. The glimpse of a column of figures, not to say an algebraic symbol or an integral sign, is likely to set off, not mathematical behavior, but a reaction of anxiety, guilt, or fear.

The teacher is usually no happier about this than the pupil. Denied the opportunity to control via the birch rod, quite at sea as to the mode of operation of the few techniques at her disposal, she spends as little time as possible on drill subjects and eagerly subscribes to philosophies of education which emphasize material of greater inherent interest. A confession of weakness is her extraordinary concern lest the child be taught something unnecessary. The repertoire to be imparted is carefully reduced to an essential minimum. In the field of spelling, for example, a great deal of time and energy has gone into discovering just those words which the young child is going to use, as if it were a crime to waste one's educational power in teaching an unnecessary word. Eventually, weakness of technique emerges in the disguise of a reformulation of the aims of education. Skills are minimized in favor of vague achievements–educating for democracy, educating the whole child, educating for life, and so on. And there the matter ends; for, unfortunately, these philosophies do not in turn suggest improvements in techniques. They offer little or no help in the design of better classroom practices.

The improvement of teaching

There would be no point in urging these objections if improvement were impossible. But the advances which have recently been made in our control of the learning process suggest a thorough revision of classroom practices and, fortunately, they tell us how the revision can be brought about. This is not, of course, the first time that the results of an experimental science have been brought to bear upon the practical problems of education. The modern classroom does not, however, offer much evidence that research in the field of learning has been respected or used. This condition is no doubt partly due to the limitations of earlier research. But it has been encouraged by a too hasty conclusion that the laboratory study of learning is inherently limited because it cannot take into account the realities of the classroom. In the light of our increasing knowledge of the learning process we should, instead, insist upon dealing with those realities and forcing a substantial change in them. Education is perhaps the most important branch of scientific technology. It deeply affects the lives of all of us. We can no longer allow the exigencies of a practical situation to suppress the tremendous improvements which are within reach. The practical situation must be changed.

There are certain questions which have to be answered in turning to the study of any new organism. What behavior is to be set up? What reinforcers are at hand? What responses are available in embarking upon a program of progressive approximation which will lead to the final form of the behavior? How can reinforcements be most efficiently scheduled to maintain the behavior in strength? These questions are all relevant in considering the problem of the child in the lower grades.

In the first place, what reinforcements are available? What does the school have in its possession which will reinforce a child? We may look first to the material to be learned, for it is possible that this will provide considerable automatic reinforcement. Children play for hours with mechanical toys, paints, scissors and paper, noise-makers, puzzles–in short, with almost anything which feeds back significant changes in the environment and is reasonably free of aversive properties. The sheer control of nature is itself reinforcing. This effect is not evident in the modern school because it is masked by the emotional responses generated by aversive control. It is true that automatic reinforcement from the manipulation of the environment is probably only a mild reinforcer and may need to be carefully husbanded, but one of the most striking principles to emerge from recent research is that the *net* amount of reinforcement is of little significance. A very slight reinforcement may be tremendously effective in controlling behavior if it is wisely used.

If the natural reinforcement inherent in the subject matter is not enough, other reinforcers must be employed. Even in school the child is occasionally permitted to do 'what he wants to do', and access to reinforcements of many sorts may be made contingent upon the more immediate consequences of the behavior to be established. Those who advocate competition as a useful social motive may wish to use the reinforcements which follow from excelling others, although there is the difficulty that in this case the reinforcement of one child is necessarily aversive to another. Next in order we might place the good will and affection of the teacher, and only when that has failed need we turn to the use of aversive stimulation.

In the second place, how are these reinforcements to be made contingent upon the desired behavior? There are two considerations here—the gradual elaboration of extremely complex patterns of behavior and the maintenance of the behavior in strength at each stage. The whole process of becoming competent in any field must be divided into a very large number of very small steps, and reinforcement must be contingent upon the accomplishment of each step. This solution to the problem of creating a complex repertoire of behavior also solves the problem of maintaining the behavior in strength. We could, of course, resort to the techniques of scheduling already developed in the study of other organisms, but in the present state of our knowledge of educational practices scheduling appears to be most effectively arranged through the design of the material to be learned. By making each successive step as small as possible, the frequency of reinforcement can be raised to a maximum, while the possibly aversive consequences of being wrong are reduced to a minimum. Other ways of designing material would yield other programs of reinforcement. Any supplementary reinforcement would probably have to be scheduled in the more traditional way.

14 The treatment of socially oriented underachievers: a case study

B. Schaeffer, A. Harris and J. M. Greenbaum

Many underachievers are socially oriented: they work only when the teacher hovers over and heckles them. They display attention-getting behaviors such as falling out of chairs, making noise, and annoying other students. They seek contact at an immature level through hitting, kissing, tapping, and pushing. They may be well liked and considered charming or they may irritate other children and be called pests. Their written school work is usually incomplete and sloppy, even though their oral work may be satisfactory. They may know the material but do not hand in assignments; they begin work enthusiastically but do not finish. Dallying behaviors such as pencil-sharpening, daydreaming, and walking around the room are common. Teachers view them as challenges and show them a great deal of coercive and affectionate attention. They keep them after class or at recess, reason with them, help them with homework, cheer them on, and allow them free access to school counselors. This attention constitutes a reward for underachievement.

The treatment program extended the work of Patterson (1965) and consisted of positively reinforcing the entire class for the underachiever's academic improvement. The socially oriented underachiever thus became the recipient of class attention, peer pressure, and social rewards, exactly as he wanted, but he could only continue as recipient if he improved academically.

Procedure

Jack, the underachiever treated, was observed (at his mother's and teacher's request) for two weeks before the program was designed and initiated. He was a short, wiry sixth-grader. His parents were divorced, and he lived with his mother. His peer group consisted of the problem children in his class; as their acknowledged leader he led forays into the school secretaries' offices after class. On the playground he was aggressive and bullying, lashing out with his fists and feet when thwarted.

His WISC [Wechsler Intelligence Scale for Children] IQ was 96 and his average grade-standing on achievement tests was 4.3, no higher than the 25th percentile on any individual test. His grades were considerably lower than average (40 to 65 per cent on spelling and arithmetic tests when the

---

Source: *Journal of School Psychology*, 1968-9, 7, 4, 70–3.

class averaged 80 to 95 per cent). He spent most of his class-time sharpening pencils, going to the restroom, flipping textbook pages, talking to and poking other students, staring off into space, and stuffing chewing gum and candy into his mouth. His attention span during written work and reading assignments was about two minutes; he did not turn in assignments. He seldom participated in class discussion, talking only to criticize or needle other students or to elaborate on irrelevant tangents. He threw temper tantrums. He could, however, be charming: the school counselor agreed to let him come talk with her whenever he felt troubled.

Treatment program

The behavior modifier taught the teacher the principles and techniques on which the program was based and worked intensively with the teacher at the start. Gradually, however, the teacher assumed control, consulting the behavior modifier only for difficult problems. The behavior modifier observed the entire program.

The program was intended to improve Jack's school work. The social environment was set up to reward appropriate academic behaviors and to punish incompatible behaviors. For every improvement in Jack's academic performance the entire class was rewarded with a minute of free time during physical education (compulsory calisthenics the class disliked) period, so that Jack was working both for himself and for the rest of the class. Every correct word over his 65 per cent spelling test average earned all the students a minute of free time. As Jack's baseline academic performance improved, the criterion for reward was changed accordingly. Improved work in arithmetic, completed assignments turned in on time, steady attention to classwork, and correct answers in class discussions were also rewarded. Jack lost free time by staring off into space, by answering poorly during class discussions, by dawdling over his schoolwork, and by being disruptive. Other students also lost the free time Jack won by excessively distracting him.

The program was initiated simply. One morning before class the teacher asked Jack privately if he would participate in a project that would help him and the rest of the class. Jack readily agreed. He was allowed only fifteen minutes to think about the program before it began, because it was thought that more time would allow him to figure out ways of using the program to manipulate the teacher and the rest of the class. The teacher then explained the program to the other students and immediately put it into effect. In a spelling test, Jack won a few minutes of free time.

Whenever Jack improved or declined academically, the teacher described the situation for Jack and the class and wrote the number of points Jack earned or lost on the board. Jack's rewards and punishments were thus immediate, and the reinforcing power of the teacher's immediate praise and approval was added to free time.

The class chose free time as a reward over early dismissal at lunch or at the end of the day. A party would be given for the class, if Jack completed an acceptable science notebook. Jack had never turned in an acceptable notebook; this delayed reward was designed to encourage the delay of gratification academic work demands.

To strengthen the effects of these contingencies, Jack was seated at an isolated front desk. This curtailed the effects of his underachieving peer group; Jack's social participation in class thus became contingent on appropriate academic behavior. Eventually Jack was reseated among a group of achievers whose behaviors he could model.

A major problem arose: the struggle for power, attention, and control between student and teacher. At the start of the program, for example, Jack was winning more than one full free physical education period per day, could afford to underachieve, and did. When he lost points, he became tense, anxious, and disruptive and lost more points, a vicious downward spiral. Changing the reward criterion so that Jack would win only one free period per day removed the problem.

At the beginning of the program, Jack asked for a spelling test because he had studied until midnight; the class enthusiastically supported his suggestion. Jack tried other attention-getting ploys: if he misspelled a word, he would ask the student grader to change his score or word. Jack also began to act the class clown, speaking out of turn, performing disruptive antics, and making inappropriate jokes. Deducting points curbed all these attention-getting behaviors. Jack's gambits were attempts to gain attention, control, and power; the teacher curbed them by punishing them consistently with the means provided by the program.

To make his eventual removal from the program possible and to strengthen appropriate academic behavior, Jack was put on a continuously thinning schedule of partial reinforcement for academic improvement. After three months, Jack was doing well enough to allow another boy, Bob, to begin the program. Since the class was satiating on free time, Jack and Bob now worked for chess pieces: when they accumulated an entire chess set, the class got a game period. To eliminate conflict between them, a bonus for good performance by the two of them was instituted.

Results

A week after Bob joined the program, Jack asked to be removed from it. He said he became overly anxious about the occasional point he lost. The teacher had wanted to remove Jack from the program, but did not know how or when; Jack's suggestion was followed. As soon as Jack was removed, his academic and social behavior declined, but a change of seat reinstated his gains, which were maintained for the remaining three months of school. (Seat changes had been tried to no avail before.)

Jack improved both academically and socially. Two months after the program began, Jack was in the upper half or quarter of his class in reading, mathematics, spelling, social studies, and Spanish; he had been in the lower half or quarter. He did class assignments and homework. He turned in an acceptable science notebook and won the class a party. In Spanish, his poorest subject, Jack formulated a question that integrated three unrelated pieces of information; the question, asked in Spanish, stumped the rest of the class, and the Spanish instructor based the day's lesson on it. Jack's IQ rose from 96 to 108 and his average achievement test grade rose from 4.3 to 5.4. Undoubtedly both improved test-taking attitudes and skills as well as actual intellectual gains were involved.

Jack's social improvement was equally dramatic. Jack was lending other students the school materials he formerly stole from them. At lunch time, instead of running away with play equipment, Jack was helping other students secure the equipment they needed. He was supportive of his fellow students when before he had been hypercritical and derisive. For example, when a classmate's report was criticized by the social gadflies of the class, Jack rationally argued away criticisms.

As a consequence, Jack became a valued social commodity. He was chosen first, second, or third as a team member during physical education instead of last or next to last; he was freely lent books, paper, pencils, and other materials. Jack's developing pro-social behaviors and his emergence as a valuable social commodity were intertwined. Before the program began, Jack was the leader of the underachievers in the class. A few weeks later he started to initiate interactions with the high achievers and leaders in the class. At first, these were rejected or only grudgingly accepted, but the achievers did initiate a few interactions and Jack persevered. Gradually Jack became a full-fledged member of the class elite. His mother described him as being happier, doing household chores and homework willingly, and no longer asking to go live with his father.

The program's results were not specific to Jack. At the start of the program there was a halo effect: the students cheered when Jack won points and hissed when he lost them. This wore off as the program became more powerful. Achievers suggested underachievers who needed the program; underachievers privately asked to be included in the program. Even after Jack was removed from the program, the class kept pressuring him, much to his dismay, to continue performing well, and, as a result of the program, the proportion of appropriate attending behaviors rose for the entire class.

Discussion

The leverage exerted by behavior modifiers can be greatly enhanced by the sharing of reinforcers throughout the social system. The free time used in this case bound the underachiever, the teacher, and the other students closely together. Because the rewards were shared, peer pressure to maintain a continual flow of rewards was generated. Integrating the entire class around the reinforcement contingency multipled the effectiveness of what might otherwise have been a weak reward. The desire of the underachiever's achieving peers for the shared reward, and their inability to influence directly how much of the reward he received, set into motion wide-ranging changes in social interaction patterns that involved the underachiever. The achieving peer group began to reinforce the underachiever for improving academically by becoming his friends, thus greatly increasing the effect of the free time reward on him and greatly increasing the leverage exerted by the behavior modifier.

A learning theory overview of the stages of the treatment process can explain why peer pressure greatly enhanced the reward's effectiveness.

1. Jack's academic improvement was spurred by positive reinforcement from the teacher.
2. Peer reinforcement for academic achievement arose because his peers shared Jack's rewards. Jack's academic improvement became a secondary reinforcer for his peers, and they plied him with social rewards for producing it.
3. Peer rewards strengthened appropriate social behaviors.
4. Increased appropriate social and academic behaviors elicited more potent peer and teacher reinforcement, and the process spiralled.
5. Appropriate academic and social behaviors, being under the same set of reinforcers, coalesced to form a unitary class of appropriate behavior.

The spiralling effects depended on Jack's academic improvement becoming a secondary reinforcer for his peers. That people and their behaviors can become secondary reinforcers is intuitively plausible. One can easily think of a person, a friend before he became a friend or a spouse before marriage, whose behaviors and opinions developed value during prolonged social interaction. The prolonged discriminative pairings of Jack's academic improvement with a positive reinforcer, free time, probably made it a potent secondary reinforcer. The academic improvement was more than a discriminative stimulus; it did more than provide free time. As a motivational force, it sustained and directed changes in Jack's social behavior that were far removed from the original reinforcement contingency.

Reference

PATTERSON, G. R. (1965) 'An application of conditioning techniques to the control of a hyperactive child.' In ULLMANN, L. P. and KRASNER, L. (eds.) *Case studies in behavior modification.* New York: Holt, Rinehart & Winston.

15 The effects of four programs of classroom intervention on the intellectual and language development of 4-year-old disadvantaged children

M. B. Karnes, J. A. Teska and A. S. Hodgins

The involvement of culturally disadvantaged children in preschool education, an experience traditionally reserved to children of more affluent parents, can now be regarded as the major educational phenomenon of the 1960s. Backed by federal funds and spurred by a sense of moral and social urgency, community groups of many different kinds launched preschool programs. Dispensing with rigid professional requirements and having little in the way of established educational theory to draw upon in meeting the special needs of these children, the programs were necessarily innovative. Even when they failed, they served to open fundamental questions about preschool education.

Interpretations of programs initiated early in this innovative and productive decade were undertaken in only the most general way. As Weikart (1967) pointed out, comparison and evaluation could not be made with conviction or precision because of the variety of programs and the lack of comparable data and followup studies. The appropriate area for experimental investigation had become the effectiveness of specific interventions rather than the question of preschool versus no preschool experience.

The study reported here was designed to evaluate the differential effects of four preschool intervention programs through pre- and post-batteries of standardized tests. The four classroom interventions were chosen on theoretical as well as practical bases to represent levels of structure along a continuum from the traditional nursery to the highly structured preschool. The nature of teacher-child interaction was considered to be the critical dimension of structure: as the specificity and intensity of this interaction increases so does the degree of structure. Two programs (Traditional and Community-Integrated) represented the less structured end of the continuum; a third (Montessori) embodied an established theory which includes much that can be identified with a child-centered or traditional approach and a methodology which incorporates considerable structure; the fourth (Experimental) was the most highly structured and emphasized learning tasks chosen from school-related curricula, especially tasks designed to enhance cognitive and language development.

Source: *American Journal of Orthopsychiatry*, 1970, *40*, 58–76.

Method

The four programs of classroom intervention

During the first year of the study, the Traditional and Experimental programs (Karnes *et al.*, 1968) enrolled a total of sixty disadvantaged children, four class units of fifteen children each. Double class units were assigned to these programs because followup studies are to continue through the early elementary grades and attrition was to be expected. During the second year of the study one class unit of sixteen was enrolled in each of the other two intervention programs (Community-Integrated and Montessori), a total of thirty-two children.[1]

Since the effects of specific classroom interventions were the concern of this study, total impact programs were not considered and variables outside the classroom were not manipulated. All children received medical examinations, but intensive medical followup was not undertaken. All parents were offered occasional opportunities to visit classrooms, but no program required intensive parental involvement or provided instruction for parents. Lunch programs were not offered at any preschool.

Each program offered morning and afternoon sessions, with the exception of the Montessori program which operated only in the morning. All children were bused to school and attended daily sessions of approximately two hours and fifteen minutes for a period of no less than seven or more than eight months. Classes were not held during the usual school holidays. Church facilities were used for classrooms in all instances except that of one Community-Integrated preschool which maintains its own school building.

A pupil-teacher ratio of 5:1 was established for the Traditional and Experimental programs, and a class unit consisted of fifteen children and three teachers. Only one qualified preschool teacher was available, and the remaining five positions were filled by college graduates experienced in working with young children. All but one was a certified teacher. An inservice training program was conducted for the Traditional teachers prior to the opening of preschool classes. Weekly inservice training sessions were held for the Experimental teachers throughout the program. The Montessori and Community-Integrated programs operated under the auspices of existing community institutions, and it was not feasible to require the modification of their teacher-pupil ratios to conform to that utilized in the other two programs. The Montessori program employed a qualified Montessori teacher and one trained teacher aide for the sixteen children, a ratio of 8:1. The pupil-teacher ratio in the Community-Integrated preschools varied from 6:1 to 10:1. Qualified preschool teachers were employed at all these centers. One center used mother aides in addition to its professional staff.

---

[1] Limited funds required a reduction from two to one class unit for these two interventions, and so the size of a class unit was increased from 15 to 16 children to compensate for attrition. It was assumed that one additional child would not alter classroom dynamics or impair comparability of groups.

The four programs of classroom intervention may be distinguished as follows:

1. Major goals of the *Traditional* nursery school program were to promote the personal, social, motor, and general language development of the children. Teachers were instructed to capitalize on opportunities for incidental and informal learning, to encourage the children to talk and to ask questions, and to stimulate their interest in the world around them. Special efforts were made to interest the children in books. Music, story, and art activities were scheduled regularly each week. Outdoor play was a part of the daily routine; indoor play focused on a doll and housekeeping center, a vehicle and block center, and a small toy center which featured puzzles, beads, puppets, books, and the like. Juice time, rest, show and tell, and the routine supervision of toileting and outdoor wraps completed the daily schedule.
2. The *Community-Integrated* program, operating at four neighborhood centers, provided a traditional nursery school experience similar to the one outlined above. These centers were licensed by the state and were sponsored by community groups, and classes were composed predominately of middle- and upper-class Caucasian children whose parents paid tuition which ranged from $18 to $40 per month. The sixteen disadvantaged children in our research unit were integrated in groups of two to four into various sessions at each of these four centers.

Socioeconomic integration was the pertinent variable rather than racial integration, which was achieved in all programs. Spontaneous verbal interactions represent critical opportunities for language development in the traditional nursery school since substantial periods of time are given to peer-initiated play. Language inadequacies of disadvantaged children might, therefore, sharply limit their progress in language development in such a setting. Central to the altered classroom dynamics in the Community-Integrated program was the presence of an advantaged-peer language model in addition to the teacher model provided in all programs. The high ratio of advantaged children assured that these children would determine the level of spontaneous verbalization. To the extent that children in a traditional nursery school acquire language from each other, the Community-Integrated program provided the optimum setting for verbal development.

3. The *Montessori* program was administered by the local society which had offered classes during the preceding three years. Staff and classroom materials met Montessori standards. The daily schedule began with a routine health check and toileting. The group then met 'on the line' for

conversation, songs, finger plays, and exercises. The following half hour was devoted to 'spontaneous choice' of approved materials: templates and stylus, cylinder blocks, dressing frames, color and weight tablets, touch boards, counting devices. The Montessori teacher noted that this class of disadvantaged children required teacher-presentation of materials more often than was typical of other classes at the school. Spontaneous choice was followed by the second period on the line devoted to musical activities, stories, and games. A 'practical life' demonstration followed: sponging exercises, fingernail care, cutting of fruits and vegetables, brushing hair and teeth. Juice time, toileting, the silence exercise, and tidying the classroom occupied the next half hour. The final ten to twenty minutes of the session were given over to playground activities or supervised short walks. Field trips received major emphasis: the fire station, the library, a shopping mall, a music store, a dog kennel, an art museum, a flower show, a farm.

The specific nature of the 'prepared environment' raised the level of structure within the Montessori classroom beyond that of the two traditional programs. The Montessori teacher did not, however, maintain the high level of specific control over the actions of the children required by the teachers in the Experimental program. Structure in the Montessori program derived not from direct teacher-child interaction but from the prescribed manner in which the child learned from the materials.

4. In the highly structured *Experimental* program, manipulative and multi-sensory materials were chosen to provide the framework for eliciting the verbal responses necessary for language development which was considered to be a critical area of deficit for disadvantaged children. The basic concepts to be taught as well as the specific learning tasks were chosen because their mastery is requisite to successful academic performance in early elementary school. Content to be learned was presented in a game format which employed manipulative materials but was structured by the teacher to require concurrent verbal responses. Teachers were instructed to accommodate their teaching strategy to the performance of the children on the initial battery of tests and to incorporate into their lesson plans the various facets of the language process embodied in the Illinois Test of Psycholinguistic Abilities.

Each class unit was divided into three groups on the basis of Binet IQ, with one teacher for each group. Groupings were flexible, however, so that children who needed extra supervision or instruction could be somewhat evenly distributed or children who did not perform according to test indications might be more appropriately placed. The daily schedule was divided into three twenty-minute structured learning periods: math concepts, language arts and reading readiness, and science-social studies. A large room where the fifteen children could gather for group activities was available; however, most of the instruction took place in relatively small cubicles off the main room. Each cubicle contained materials appropriate to one of the three content areas, and each teacher moved from one cubicle to another with her group of five children.

Since the teacher-child relationship is of primary importance in securing motivation and in providing opportunities for the reinforcement of learning, each group remained with the same teacher for the three structured periods, for juice, and for field trips. The low teaching ratio allowed for differentiation of instruction to provide a high success ratio for each child. Immediate correction of incorrect responses (often through repetition of model sentences or duplicate layouts of small, manipulative materials) and reinforcement of appropriate responses (usually through praise) assured the children of their competencies in handling curricular requirements and enhanced their intrinsic motivation to learn. Frequent review extended content previously presented and provided opportunities to use further the vocabulary and sentence structures which had been taught.

Children formed their own peer groupings during the music period and during a brief period of directed play which stressed visual-motor activities such as puzzles, blocks, clay, and pounding sets. No use was made of outdoor play equipment or traditional preschool toys such as dolls, toy appliances, cars, or trucks. Concepts taught during the structured periods were reinforced during directed play and especially during the music period. For example, when body parts were introduced in science or counting in math, these concepts were stressed in songs and rhythmic activities during music.

The general goals of the social studies and science curriculum were to teach useful vocabulary, to develop skills of classification, to provide simple experiences in developing sensory discriminations and in observing natural phenomena. The curriculum began with a unit on body awareness and self-concept developed through the use of body exercises, songs, precut unassembled figures, and body outlines of the children. A unit on family members and immediate home environment followed, using integrated pictures, rubber play people, and family puppets; clothing cut from catalogs and sorted according to body parts, family member, or season; furniture items cut from catalogs and sorted according to type or appropriate room; go-together pictures such as a hand and a mitten, a chair and a table. A kitchen science unit, through the demonstration of simple scientific principles, provided opportunities for careful observation and verbalization of what had been seen, heard, tasted, or touched. Basic vocabulary included *melt, boil,* and *freeze; dry* and *wet*; relative temperature words such as *cool, warm,* and *hot; dissolve;* taste words such as *sweet, sour,* and *salty.* Additional units in this curriculum were germination of seeds and plant growth, farm and wild animals, fruits, vegetables, community buildings and workers, vehicles, weather, seasons, and time sense.

Objectives of the math curriculum involved the development of basic number concepts, appropriate manipulative skills, and a useful vocabulary. The general areas included the identification of five geometric shapes;

one-to-one matching and its relationship to copying patterns, matching quantity, and establishing sets and verifying their equivalency; dimensional terms and seriation; counting as a functional concept; the introduction of numerals as visual symbols; and beginning addition and subtraction with manipulative objects such as popsicle sticks, bottle caps, and peg boards.

Multiple copies of inexpensive books were the most important instructional material in the language arts and reading readiness curriculum. As the teacher read, each child held his own copy of the book; he learned to hold the book right-side-up, to turn the pages singly and in sequence, to associate the pictures with the story being read, to develop left-to-right progression, and to associate the printed symbol with meaning. In addition, the small group storytime provided opportunities for reinforcing and elaborating upon vocabulary previously taught; for both short- and long-range memory activities; for sequencing events to show cause and effect and time relationships; for making inferences and, on occasion, divergent responses. Finally, as the story was read, the child heard acceptable syntactical models and the familiar constructs of the language. He absorbed the rhythms and stresses of standard, informal English. This curriculum also included activities which developed visual-motor coordination and which emphasized the rather fine visual and auditory discriminations requisite for reading readiness.

Language development received major emphasis throughout the day, especially during the three structured periods. Verbalizations in conjunction with the manipulation of concrete materials were considered to be the most effective means of establishing new language responses. The game format (card packs, lotto games, models and miniatures, sorting, matching, and classifying games) created situations where verbal responses could be made repeatedly in a productive, meaningful context without resorting to rote repetition; often the child could visually and motorically assess the correctness of his thinking before he made an appropriate verbalization. If the child was unable to make a verbal response, the teacher supplied an appropriate model; when he began to initiate such responses, the teacher had the opportunity to correct, modify, and expand his verbalizations.

Selection of subjects

The subjects for this study were selected from the preschool population of the economically depressed neighborhoods of Champaign-Urbana, a community of 100,000 in central Illinois. Families judged by public aid and school authorities to be economically and educationally deprived were canvassed for children who had no previous preschool experience and who would be four years old before the first of December, an age appropriate for enrollment in public kindergarten the following year. This age criterion was established so that followup evaluations could be more efficiently coordinated with the public schools. The mean chronological ages of the children in the four groups at the time of the initial Stanford-Binet ranged from 49 to 52 months. A home interviewer determined final eligibility for the program after she had completed a detailed family history. In addition, interviewers canvassed certain acutely disadvantaged sections of the city to locate children who might be new to the community or otherwise unknown to the referring agencies.

The 1960 Stanford-Binet Intelligence Scale was administered to eligible children who were then stratified on the basis of their intelligence quotients into three groups: IQ scores 100 and above, 90 through 99, and 70 through 89 (two children with IQs below 70–69 and 67– were included). Typically, each class unit of fifteen included five children who had scored above 100, five who had scored between 90 and 99, and five who had scored between 70 and 89. Mean IQs were then computed for the three strata and for each class unit. These means were evaluated for comparability between class units as a whole and for strata between classes. These strata ensured a balanced range of intelligence scores in each class unit and provided an opportunity to evaluate the effectiveness of the various programs on children from different ability groups. The mean IQ (approximately 95) of the children placed in the classes was, of course, higher than the mean of the children screened.

Class units were examined to assure comparability of sex and race. When necessary, substitutions were made between classes to maintain an approximate ratio of 67 per cent Negro children and 33 per cent Caucasian children and a ratio of approximately 50 per cent male and 50 per cent female children. Finally, each class unit was randomly assigned to a particular intervention program.

Evaluation procedures

Since the intent of this study was to evaluate the effectiveness of four classroom interventions upon the overall school readiness of disadvantaged children, pre- and post-evaluations were made in the following areas:

1. Intellectual functioning as measured by the 1960 Stanford-Binet Individual Intelligence Scale, Form L-M.
2. Language development as measured by the Illinois Test of Psycholinguistic Abilities, experimental edition, 1961.
3. Vocabulary comprehension as measured by the Peabody Picture Vocabulary Test.

In addition, the Frostig Developmental Test of Visual Perception was administered at the time of the post-battery. Qualified psychological examiners administered the tests at a school site and were not informed of the program assignment of the children.

Statistical procedure

Statistical treatment of the data employed a repeated measures analyses of variance followed by Newman-Keuls tests at the 0.05 level for significance of differences among means. For brevity of presentation, statistical tables are not included.

Results and discussion

Intellectual functioning

Clearly, the Experimental group made greater gains in intellectual functioning as measured by the Stanford-Binet than the other three groups (Table 1). Less than three Binet IQ points separated the four groups on Test 1, a nonsignificant difference. On Test 2 the Experimental group performed significantly higher than the other three groups; there were no significant differences among the others. Only the Community-Integrated group failed to score significantly higher on Test 2 than on Test 1.

Table 1 *Mean Stanford-Binet IQ*

Group	*Test 1*	*Difference*	*Test 2*
Traditional	94.5	8.1	102.6
Community-Integrated	93.3	5.1	98.4
Montessori	94.1	5.5	99.6
Experimental	96.0	14.3	110.3

The initial IQ scores of one-third of the children in each unit were in the high strata, 100 and above. On Test 2, 93 per cent of the children in the Experimental group fell in this strata while only 31 to 56 per cent of the children in the other groups earned such scores. The two children in the Experimental group who were not in the high strata on Test 2 scored 96 and 98 and had initially scored in the low (70–89) strata. Not only did significant numbers of children in the other three programs fail to score 100 at the time of Test 2, but 14 to 31 per cent fell in the low strata.

The distribution of gains (Table 2) emphasizes the magnitude and the consistency of the gains achieved by the children in the Experimental program. Seventy-four per cent of the children in the Experimental group made gains of ten or more points; only 30 to 39 per cent of the children in the other three groups made similar gains. No child in the Experimental group failed to make a gain. (The one child who fell in the 0–4 range gained four points.) Twenty-one to twenty-four per cent of the children in the other three groups scored lower on Test 2 than on Test 1.

Table 2 *Distribution of IQ gains by intervention*

Gain in IQ points	*Group*							
	*Traditional N=28**		*Community-Integrated N=16*		*Montessori N=16*		*Experimental N=27**	
	%	*N*	*%*	*N*	*%*	*N*	*%*	*N*
25 to 29	0	(0)	12	(2)	0	(0)	7	(2)
20 to 24	7	(2)	6	(1)	0	(0)	15	(4)
15 to 19	21	(6)	0	(0)	12	(2)	22	(6)
10 to 14	11	(3)	12	(2)	19	(3)	30	(8)
5 to 9	29	(8)	6	(1)	31	(5)	22	(6)
0 to 4	11	(3)	38	(6)	12	(2)	4	(1)
– 1 to – 5	21	(6)	12	(2)	12	(2)	0	(0)
– 6 to –10	0	(0)	6	(1)	12	(2)	0	(0)
–11 to –15	0	(0)	0	(0)	0	(0)	0	(0)
–16 to –20	0	(0)	6	(1)	0	(0)	0	(0)

*The Traditional and Experimental groups had an initial *N* of 30 each. Five children withdrew from the program before the posttest battery and no data for these children is included in this study.

The distribution data reflect one of the most important findings of this study: the Experimental program had a positive effect on the IQ score of *every* child in the group. One must assume that the highly structured Experimental program provided unique opportunities for enhancing the level of intellectual functioning as well as effecting this change with remarkable consistency. Structure predicates active involvement of teacher and child. Through manipulative experience, the child moved to physical mastery of a concept and was required by the teacher to make appropriate verbalizations. Moving from structured, physical involvement within a meaningful, productive context to independent, conceptual verbalizations is fundamental to every aspect of intellectual functioning. The teacher in this setting monitored the child's manipulative performances and assessed the adequacy of his verbal responses so that she could alter the learning situation appropriately. It was the function of the teacher to provide sufficient repetition to establish new verbal responses and to alter the learning task to encompass further cognitive and verbal complexities. In the other three programs a variety of learning experiences was made available to the children, but their involvement in specific experiences was not required and concurrent verbal responses were not insisted upon. Individual children in the other three programs did indeed make gains greater than the highest gain made by any child in the Experimental program. However, the number of children who made excellent gains was overshadowed by the percentage who made minimal gains and regressions.

The distribution of gains in the Community-Integrated group merits comment, particularly since this was the only intervention in which disadvantaged children were integrated with advantaged children. Sixty-two per cent of the disadvantaged children in this group made minimal gains (0–4 points) or lost in IQ. At the other extreme, three children (18 per cent) made large gains and were essentially responsible for the mean gain of the entire group.[2] Since the Community-Integrated and Traditional programs provided similar experiences, one would expect a parallel distribution of gains. Since this was not the case, it is reasonable to assume that the atypical distribution was produced by altered classroom dynamics: the presence of the advantaged children as the majority group.

The disadvantaged children in the Community-Integrated setting remained somewhat uninvolved in certain critical aspects of the program, particularly in the quasi-structured areas of the daily schedule which involved language: show and tell, music, storytime, and discussions at circle time which included the calendar and easel pictures. The disadvantaged children typically sat in the fringe area at storytime and were observed to remain aloof or to attend absentmindedly. Observers recorded: 'Played with the shadows made from his fingers throughout the story.' 'Fiddled with the folds in the window curtains while the story was read.' 'Sat with his rug on his head during presentation of the easel picture.' During music they were observed 'making the motions but not singing the words'.

Neither the teachers of the Community-Integrated group nor the advantaged children rejected the disadvantaged children. In fact, these teachers were seen as giving the disadvantaged children more than their share of attention. It is fair to add, however, that this attention was sometimes given to the acting out or aggressive child rather than to the aloof or uninvolved child.

The disadvantaged children functioned at least adequately in areas of the daily schedule which emphasized large motor development and social skills and sought these areas for themselves. The block and vehicle center was popular with the boys as was the doll and housekeeping center with the girls. Even in these areas, however, the disadvantaged child's verbal functioning was minimal. One girl, for instance, played house happily and well for an hour with a boy from the advantaged segment of the class. He spontaneously verbalized his play activity; she, however, cleaned the house, made breakfast, washed the dishes, and even packed his lunch without speaking. It is reasonable to conclude that the disadvantaged children sought situations in which they felt competent and avoided those in which they felt inadequate.

Since the spontaneous speech of the advantaged children in this group was well developed, the silence of the disadvantaged children went relatively unmarked by the Community-Integrated teachers. The songs were well sung, the easel pictures and the calendar were developed in great verbal detail, but the dominant group, the advantaged children, were responsible for the success of these activities. In the Traditional group, however, these quasi-structured and basically verbal activities had to be sustained through the participation of disadvantaged children. In this setting the teachers were keenly aware of the language inadequacies of their group. The children could not avoid these situations or remain uninvolved because there were no other children to replace them in sustaining the activity. Implicit in the equipment, in the activities which make up the daily schedule, and in the philosophy of the traditional nursery school is the assumed middle-class background of the children: four-year-old children enjoy listening to stories; four-year-old children spontaneously verbalize about their environment; four-year-old children talk freely during their play. The Community-Integrated teachers could operate effectively within these assumptions because the majority of their children met these expectations; the teachers in the Traditional program necessarily accommodated the operation of their program to the background and performance of the children.

---

[2] No obvious characteristic of these three children or their group placement seems to have been related to their high gains. A particular community preschool did not account for these gains since each of the three attended a different preschool. Their racial distribution was the same as that of their intervention unit, two Negro and one Caucasian. There were two females and one male, and there was one child from each of the three intelligence strata.

Language development

The language performance of the Experimental group is a matter of real importance since its program focused on specific verbal skills and learning activities emphasized the co-relation of cognitive and language development. Teachers incorporated the various facets of the language process embodied in the Illinois Test of Psycholinguistic Abilities (ITPA) into their lesson planning and adjusted their teaching strategy to initial ITPA profiles of individual children. Opportunities for language development in the Traditional program occurred in more general ways, particularly in quasi-structured activities such as show and tell, circle time, and music and in spontaneous situations arising from play. Peer language models should have been more influential in the Traditional group than in the Experimental group which had fewer opportunities for spontaneous language. Situations affecting language development in the Community-Integrated group were similar to those in the Traditional group but were obviously altered by the introduction of peer language models from an advantaged segment of the population. Finally, the Montessori program focused on motor-sensory learning as the basic mode in which conceptual and linguistic abilities occur, following the pattern of the child's sensorial development.

Comparisons of ITPA language age scores are confounded by slight differences in initial mean chronological ages of the groups and slight variations in test intervals. To compensate for these variations and to relate language age to chronological age, a difference score was computed by subtracting a child's chronological age at the time of testing from his language age. A child who was 48 months old and earned a language age score of 40 months on a given subtest received a difference score of −8 months; that is, he had a deficit of eight months in that area. It must be kept in mind that a difference score gain of five months between Test 1 and Test 2 would represent thirteen months change in language age score: five months reduction of deficit plus eight months gain required by the interval between tests. Children who scored below the norms provided in the Examiner's Manual were arbitrarily assigned the lowest language age score of that subtest.

Since the intelligence strata design of this study produced an inflated Binet IQ mean compared to the mean of the disadvantaged children screened, it is reasonable to assume that this procedure also produced inflated ITPA means; and it is not surprising to discover that these means do not reflect deficits in all areas of the ITPA. The four intervention groups were not matched on ITPA performance, and Test 1 differences occurred which in two instances reached statistical significance.

The ability to express oneself verbally is the common requisite for successful performance on the three subtests in which the disadvantaged children in this study evidenced large initial deficits (Table 3):

Table 3 *Illinois Test of Psycholinguistic Abilities: difference score means in months: subtests of large initial deficit*

Group	*Test 1*	*Difference*	*Test 2*
	Vocal encoding test		
Traditional*	−14.8	10.0	−4.8
Community-Integrated	−8.2	−4.7	−12.9
Montessori	−11.8	−2.2	−14.0
Experimental	−13.6	11.2	−2.4
	Auditory-vocal automatic test		
Traditional	−8.3	3.9	−4.4
Community-Integrated	−14.1	−1.0	−15.1
Montessori	−8.4	−10.2	−18.6
Experimental	−10.8	10.7	−0.1
	Auditory-vocal association test		
Traditional	−5.9	1.9	−4.0
Community-Integrated	−11.1	1.5	−9.6
Montessori	−7.6	0.4	−7.2
Experimental	−5.4	6.7	1.3

*Two of the original 28 children in this group were by error not given a complete initial test battery, and no ITPA data are presented for these two children.

1. The Traditional and Experimental groups made significant progress on the *Vocal Encoding* test, and the Experimental group essentially eliminated its deficit. The regressions of the Community-Integrated and Montessori groups were nonsignificant.
2. Only the Experimental group made significant progress on the *Auditory-Vocal Automatic* test and eliminated its deficit. The Traditional and Community-Integrated groups remained relatively unchanged, while the Montessori group regressed significantly.
3. On the *Auditory-Vocal Association* test only the Experimental group made significant progress, and for the third time eliminated a major initial deficit. The performances of the other three groups remained basically unaltered.

The Community-Integrated and Montessori groups demonstrated a static or somewhat regressive pattern on these three important subtests. Observers who visited the Montessori classroom noted the lack of verbal expressive experiences: Children were engaged with motor-sensory materials for remarkably long periods of time, but silence prevailed. More opportunities for such experiences existed in the Community-Integrated program; however, as noted previously, the disadvantaged children tended to withdraw from these situations. Children in the Traditional group, on the other hand, did relatively well on these three subtests. When children participated in verbal expressive activities, as the dynamics of homogeneity required, they did indeed make gains. The superior performance of the Experimental group was clearly demonstrated. It seems reasonable to conclude that physical involvement alone, no matter how well programmed, sustained, and orderly, did not foster the verbal expressive development of disadvantaged children. Neither did spontaneous or relatively unstructured opportunities for such development prove adequate. When, however, activities were structured to engage the child in physical manipulation while concurrent, meaningful verbalizations were elicited, verbal expressive abilities dramatically improved.

On three subtests the four groups demonstrated relatively small initial deficits (Table 4):

1. There were no significant differences among the performances of the four groups on the *Motor Encoding* test. With the exception of the Community-Integrated group, all groups made modest progress.
2. Because initial deficits varied significantly on the *Visual-Motor Sequencing* test, the relation between program variables and performance is unclear. Only the Traditional group performed significantly higher on Test 2 than it had on Test 1. The Experimental group made a relatively large gain (eight months) and eliminated its initial deficit. Only the Montessori group failed to perform at or near its chronological age on Test 2, and, in fact, regressed, scoring six months below its mean chronological age.

Table 4 *Illinois Test of Psycholinguistic Abilities: difference score means in months: subtests of small initial deficit*

Group	*Test 1*	*Difference*	*Test 2*
	Motor encoding test		
Traditional	−2.1	4.2	2.1
Community-Integrated	−3.7	0.3	−3.4
Montessori	−5.2	4.0	−1.2
Experimental	−6.6	6.7	0.1
	Visual-motor sequencing test		
Traditional	−11.1	8.9	−2.2
Community-Integrated	−1.9	2.3	0.4
Montessori	−1.1	−4.7	−5.8
Experimental	−6.4	7.8	1.4
	Auditory decoding test		
Traditional	−4.0	5.3	1.3
Community-Integrated	−5.9	6.7	0.8
Montessori	−0.5	−0.9	−1.4
Experimental	1.9	0.5	2.4

3. The four groups did not make significantly different progress on the *Auditory Decoding* test, and all Test 2 performances were nondeficit. The Traditional and Community-Integrated groups did, however, show the largest gains and eliminated modest initial deficits.

On three subtests the four groups demonstrated little or no initial deficit (Table 5):

1. No group made significant progress on the *Auditory-Vocal Sequencing* Test. The Test 2 performances of the Traditional and Experimental groups were significantly higher than those of the other two groups; however, the regressions of the Montessori and Community-Integrated groups (approximately six months) very nearly reached statistical significance and essentially accounted for these statistical differences. The regressive performance of the Community-Integrated and Montessori groups may be related to their relatively poor Test 2 performances on all tests requiring verbal interaction with the examiner (Vocal Encoding, Auditory-Vocal Automatic, and Auditory-Vocal Association).
2. Only the Experimental group made significant progress on the *Visual Decoding* test, scoring well above its chronological age at the time of Test 2. The naming and labelling activities prevalent in the three curricula, especially activities which emphasized matching, sorting, and classification,

Table 5 *Illinois Test of Psycholinguistic Abilities: difference score means in months: subtests of minimal or no initial deficit*

Group	*Test 1*	*Difference*	*Test 2*
Auditory-vocal sequencing test			
Traditional	2.3	0.1	2.4
Community-Integrated	−0.6	− 6.5	− 7.1
Montessori	−0.7	− 5.6	− 6.3
Experimental	0.8	5.3	6.1
Visual decoding test			
Traditional	−1.5	4.5	3.0
Community-Integrated	−2.8	7.4	4.6
Montessori	−0.1	9.5	9.4
Experimental	−0.2	11.9	11.7
Visual-motor association test			
Traditional	−1.8	0.6	− 1.2
Community-Integrated	−0.2	7.3	7.1
Montessori	0.4	− 1.7	− 1.3
Experimental	9.9	0.7	10.6

may have helped the Experimental group to make this gain. Such activities were scheduled to provide opportunities for vocabulary development and for more precise, expanded, and flexible use of language; however, the *method* of presentation relied heavily on visual materials. All groups did well in this area, however, and the gain (ten months) of the Montessori group approximated the gain (twelve months) of the Experimental group. The use of manipulative and visually self-corrective materials in the Montessori program may have helped those children to perform well in this area. These activities, however, did not similarly accelerate the performance of the Montessori group on the Visual-Motor Association test. At any rate, the Montessori group made substantial progress on the Visual Decoding test in marked contrast to its generally static or regressive ITPA performance.
3. No group made significant progress on the *Visual-Motor Association* test. The error term obtained in the analysis of variance was substantially larger for this subtest than for any other ITPA subtest and mitigated against the achievement of significant differences.

Only the Traditional and Experimental groups made significant progress on the *ITPA Total* (Table 6). There were no significant initial differences among the four groups. The Test 2 performance of the Experimental group was significantly higher than that of the other three groups; the Test 2 performance of the Traditional group was significantly higher than that of

Table 6 *Illinois Test of Psycholinguistic Abilities: difference score means in months: total*

Group	Test 1	Difference	Test 2
Traditional	−5.3	4.3	−1.0
Community-Integrated	−6.3	1.1	−5.2
Montessori	−3.9	−1.4	−5.3
Experimental	−3.3	6.2	2.9

the Community-Integrated and Montessori groups whose Test 2 performances did not differ from each other. This statistical ranking generally reflects the data obtained from the nine subtests. On four of the nine subtests the Test 2 performances of the Experimental group were significantly higher than its Test 1 performances. The Traditional group achieved statistical significance in two such instances. In no instance were the performances of the Community-Integrated and Montessori groups significantly higher on Test 2 than on Test 1, and in one instance the performance of the Montessori group was significantly lower on Test 2.

The Experimental group moved in a positive direction on all subtests. The range of this upward movement was from 0.5 to 11.9 months, and on seven of the nine subtests this gain exceeded five months.[3] The Traditional group also moved in a positive direction on all subtests. The range was from 0.1 to 10.0 months, and on three subtests this gain exceeded five months. The Community-Integrated and Montessori groups did not consistently move in a positive direction, reflecting a pattern of more moderate gains and losses. On six of the nine subtests the Community-Integrated group made gains ranging from 0.3 to 7.4 months; in three instances these gains exceeded five months. On three subtests, however, the Community-Integrated group demonstrated regressions which in one instance exceeded five months. The Montessori group moved in a positive direction on three of the nine subtest with gains ranging from 0.4 to 9.5 months. Only one gain exceeded five months. On six subtests, however, the Montessori group demonstrated regressions which in two instances exceeded five months.

Vocabulary comprehension

The relative progress in vocabulary comprehension of the four groups approximated the pattern shown by ITPA and Binet data. The Traditional and Experimental groups made significant gains while the Community-

---

[3] Five months gain or loss was chosen as a descriptive evaluation point since in half of the instances language age difference scores for the four groups were altered to that extent.

Table 7 *Peabody Picture Vocabulary Test: Mean IQ*

Group	*Test 1*	*Difference*	*Test 2*
Traditional *	80.2	12.4	92.6
Community-Integrated	81.4	4.2	85.6
Montessori	83.3	4.0	87.3
Experimental	85.8	10.3	96.1

* Two of the original 28 children in this group were by error not given a complete initial test battery, and no PVVT data are presented for these two children.

Integrated and Montessori groups made small, nonsignificant gains (Table 7). Since this instrument assesses vocabulary comprehension through motor response, the verbal expressive strengths demonstrated by the Experimental children on the ITPA were not involved.

Visual perception

The Frostig Developmental Test of Visual Perception was given to the four groups at the time of the second test battery. The mean perceptual quotient of the Experimental group (99.1) was significantly higher than those of the other three groups (from 80.4 to 85.5) which did not differ significantly.

Frostig (1964) suggests that children whose scores fall in the lowest quartile (a perceptual quotient of 90 or below) will experience difficulty in school adjustment and recommends remedial training. The Test 1 scores of 95 per cent of the children in the Traditional program and 78 per cent of the children in the Experimental program fell in the lowest quartile. Test 1 data were not obtained for the Montessori and Community-Integrated groups; however, Test 1 data for four subsequent research class units established according to the procedures outlined in this paper are consistent with the scores of the Traditional and Experimental groups. In these four classes the percentages of children initially scoring in the lowest quartile ranged from 81 to 88, and it is reasonable to assume that the Montessori and Community-Integrated groups would have included a similar percentage of children in this low range. On Test 2, 75 per cent (twelve) of the children in the Montessori group, 77 per cent (twenty) of the children in the Traditional group, and 81 per cent (thirteen) of the children in the Community-Integrated group scored in the lowest quartile. Only 26 per cent (seven) of the children in the Experimental group obtained such scores.

The visual-motor activities included in the Traditional and Community-Integrated programs were relatively unstructured and global in nature (playground activities, art projects, puzzles, and other small motor toys). The children in the Montessori group did not do well on this test in spite of the motor-sensory emphasis (templates, form boxes, kinesthetic alphabet, sensory materials related to shape, weight, and size). The Experimental curriculum included learning activities designed to contribute to the development of visual perceptual skills: the unit on geometric shapes in the mathematics curriculum, a sequence of cutting lessons, dot-to-dot exercises in a large, uncluttered format, matching exercises, pasting exercises which emphasized figure-ground, and pencil/crayon work in general. Frostig remedial materials were used during the last six weeks of the program.

Summary of results

Clearly, the Experimental group made greater gains in *intellectual functioning* (Stanford-Binet) than the other three groups. Children in this group consistently made substantial gains, and a remarkable percentage moved into the high IQ stratum, 100 and above.

On the initial assessment of *language development* (ITPA), the children in this study were most deficit on three subtests related to verbal expressive abilities. The Test 2 performance of the Experimental group on these subtests was significantly superior to its Test 1 performance, and substantial deficits were eliminated. The Traditional group did well in this area, obtaining statistically significant gains on one subtest. The performances of the Community-Integrated and Montessori groups were static at best, and their substantial deficits remained at the time of Test 2. On one subtest, in fact, the Montessori group showed a significant regression. The magnitude of the gains of the Experimental group on the nine subtests of the ITPA and the consistency with which it made these gains resulted in an essentially nondeficit Test 2 performance. The Traditional group made consistent but more modest gains. The Community-Integrated and Montessori groups, on the other hand, generally made smaller and less consistent gains. The movement of the Montessori group tended to be regressive while that of the Community-Integrated group was more nearly static.

The Montessori and Community-Integrated programs were least effective in promoting *vocabulary comprehension* (PPVT). The other two programs were significantly superior and equally effective in this area. The Test 2 performance of the Experimental group in *visual perception* (Frostig) was significantly superior to those of the other three groups.

The performances of the four intervention groups on the battery of tests chosen to evaluate various aspects of school readiness differentiated in a relatively consistent order. The Experimental and Traditional groups moved in a positive direction on all instruments; however, the gains of the Experimental group were generally larger and achieved with greater consistency. In contrast, the Community-Integrated and Montessori groups showed smaller and less consistent gains, remaining in a relatively static position.

Conclusions

Since the four intervention programs were chosen to represent points along a continuum of structure, one might assume that the results would order themselves along this continuum to the extent that structure is a valid dimension in effecting change. Such was not the case. The children in the Experimental program (high on the structure continuum) showed the greatest gains. Those who participated in the Traditional program (low on the structure continuum) showed modest gains. Children in the Community-Integrated program (also low on the structure continuum) made relatively little progress. Those who participated in the Montessori program (midway on the structure continuum) showed the least progress.

The failure of the Montessori children to demonstrate appreciable progress seems to invalidate the notion that structure is essential for maximum progress of the disadvantaged child. This paradox may be more apparent than real. The Montessori program provided a high degree of structure in terms of careful planning for the kinds of motor-sensory activity appropriate to the development of an adequate base from which language and cognitive skills arise. Such provisions in terms of teacher planning and available materials may be considered comparable to the activities used to elicit verbal responses in the Experimental program. The resolution of the paradox may lie in the nature of the teacher-child interaction obtaining in the two programs. The distinguishing characteristic of the Experimental program was the tying of verbalizations to motor-sensory performance. As the child visually and motorically assessed the correctness of his thinking, he was required to make appropriate verbalizations at every stage of task involvement. The teacher began by supplying verbal models, and the repetition involved in the game format helped the child establish these verbalizations. When the child was able to initiate such responses, the teacher helped him modify and expand them. In contrast, the absence of language concomitant with performance characterized the Montessori classroom. The teacher provided a 'prepared environment'; she did not, however, systematically engage the child in verbalizations around these experiences or require such verbalizations as part of the definition of productive involvement. The failure of the Montessori program to unite verbal and motor-sensory experience resulted, at least during the intervention interval, in somewhat regressive language behavior. Placing disadvantaged children in a setting which does not require that they use the verbal symbols they have and which accepts behavior which does not involve the acquisition of new verbal skills resulted in the poorest performance of the four intervention groups. Structured emphasis on motor-sensory development without similar concern for verbal development programmatically moves in the wrong direction for the disadvantaged child.

Finally, arbitrary elements in the research design of this study may have inhibited the effectiveness of the Montessori program more than the other three interventions. First, the age criterion used in this study did not coincide with the optimum and earlier age for intervention advocated by Montessori. Second, a program interval of such duration is inconsistent with the relatively long-term Montessori developmental theory. Third, proponents of Montessori theory might well feel that the establishment of an adequate motor-sensory base, even during so truncated an interval and at so relatively late an age, will enhance language and cognitive development at a time beyond the scope of this report. Followup studies of these children may support such a point of view.

The expectation that children in the Community-Integrated group would show progress equal to or greater than that of the children in the Traditional group was not substantiated. The integration of small numbers ($N = 2$ to 4) of disadvantaged children into classes of middle- and upper-class children was intended to alter certain aspects of the learning situation. Central to the altered classroom dynamics was the presence of an advantaged peer language model in addition to the teacher model provided in all programs. Theoretically, spontaneous verbal interactions in both traditional programs represented critical opportunities for language development and were weighted in favor of the Community-Integrated participants. The disadvantaged children in the Community-Integrated program, however, failed to incorporate the language model of their advantaged peers because they did not reciprocate in verbal interactions at any significant level. They not only failed to interact verbally in peer-initiated play situations but tended to withdraw from quasi-structured, teacher-directed activities and thus sharply limited the progress they were to make. The homogeneity of the Traditional group, on the other hand, required these children to respond verbally during such activities. Their teachers necessarily accommodated these activities to the verbal level of the children and gradually developed more aceptable and extended responses. The progress in verbal expressive ability made by the children in the Traditional program reflects this accommodation.

The failure of socioeconomic integration in this instance cannot be viewed as evidence that disadvantaged children do not derive benefits from integration with their advantaged peers. It is altogether possible that the ratio of socioeconomic integration used in this study mitigated against change. A more nearly equitable ratio between the two socioeconomic groups might have created a setting in which the disadvantaged children participated more freely and were involved in verbal interactions more frequently. Then, too, it is possible that the impact of a peer language model cannot be felt during so short an interval or that four may not be the most appropriate age for deriving maximum benefits from a peer language model.

The progress made by the children in the Traditional program must be viewed against the superior performance of the children in the Experimental program. The magnitude and consistency of their gains in intellectual functioning (Binet IQ) clearly endorse the importance of providing a setting in which the child moves to physical mastery of a concept while he is required to make appropriate and increasingly complex verbalizations. In addition, verbal expressive abilities (the most severely deficit subtest of the ITPA) were dramatically accelerated by such structure. Because cognitive development at more complex levels hinges upon the existence of verbal expressive abilities, the language deficit of the disadvantaged child is of critical importance. It is precisely by connecting cognitive development and verbal expression through structured learning situations that the Experimental program demonstrated its greatest strength.

References

FROSTIG, M. (1964) *Administration and Scoring Manual for the Marianne Frostig Developmental Test of Visual Perception.* Palo Alto, Calif: Consulting Psychologists Press.

KARNES, M., WOLLERSHEIM, J., STONEBURNER, R., HODGINS, A., and TESKA, J. (1968) 'An evaluation of two preschool programs for disadvantaged children: a traditional and a highly structured experimental preschool', *Excep. Child., 34,* 667–76.

WEIKART, D. (1967) 'Preschool programs: preliminary findings', *J. spec. Educ., 1,* 163–81.

Section Four
The course of development

Introduction

The four papers in this section are all concerned with 'ages and stages' in development. They each focus on ways in which the course of an individual's development goes through different phases, in which different things are happening in the growth of the child's view of himself and the world.

Mussen and Jones start from the fact that children do not all develop at the same rate. Their study concentrates on adolescence, a period in development in which a whole variety of changes occur–some due to physical maturation, others to social structures and forces, but all interlinked in a complex of causes and effects. The adolescent who is a 'late-maturer' might be expected to have a very different status from one of the same age who is an 'early-maturer'. By using the Thematic Apperception Test (basically a set of ambiguous drawings of social situations), the authors of this paper explore the effects that different rates of maturation have on adolescents' self-concepts. Although the results are complex, one thing stands out clearly: if your body matures early, then your view of yourself will be a stronger, more positive one.

Olver and Rigney Hornsby's paper is centred on a major area of cognitive development: the ability to conceptualize or, as they put it, to 'group as equivalent'. They show how the bases that children use to group objects together can take a number of different forms, and that a clear progression is evident from such approaches as 'edge-matching' and 'key rings' to the more mature use of superordinate categories by older children. For example, a young child may group a potato, meat and milk together, saying, 'potato and meat are served together and you get milk from cows' (an 'edge-matching'). However, an older child is more likely to group these three together as being 'all things you eat'. Of particular interest is Olver and Hornsby's demonstration that young children appear to operate predominantly on the basis of perceptual features of the environment.

The third paper in this section is also concerned with intellectual aspects of development. This is a classic extract from one of the many books written by Piaget and his co-workers, which sets up some provocative hypotheses about how children come to a full understanding of 'number'. The basic thesis proposed is that the young child is 'perceptive, but not thoughtful', that what he lacks is a strong conceptual structure to mediate moment-to-moment perceptions. Thus, he will often assert that a row of counters 'has more in' because it is longer, not because it contains a greater number of counters.

Bryant's paper takes a detailed and critical look at the suppositions made by Piaget's theory, and describes some ingenious experiments which suggest that the young child is not quite as 'thoughtless' as Piaget would have us believe. Bryant accepts Piaget's observations, but invokes a radically different interpretation. The differences between these two interpretations are important for our understanding not only of the stages of cognitive development but also of the ways in which teachers can best arrange for their pupils to learn.

John Oates

16 Self-conceptions, motivations and interpersonal attitudes of late- and early-maturing boys

P. H. Mussen and M. C. Jones

While many intensive case studies show that personal and social adjustment during adolescence may be profoundly influenced by rate of physical maturation, there is a scarcity of systematic data on the relationship between the adolescent's physical status and his underlying motivations, self-conceptions and interpersonal attitudes. There is, however, a small body of evidence which demonstrates that greater physical maturity is associated with greater maturity of interest among girls (Stone and Baker, 1939) and that early-maturing boys differ from their late-maturing peers in both overt behavior and reputational status. In one study (Jones and Bayley, 1950) in which a staff of trained observers assessed a large group of adolescents on a number of personality variables, boys who were consistently retarded in physical development were rated lower than those who were consistently accelerated, in physical attractiveness, grooming, and matter-of-factness; and higher in sociability, social initiative (often of a childish, attention-getting sort), and eagerness. Reputation Test (Tryon, 1939) data indicated that classmates regarded the late-maturing boys as more attention-getting, more restless, more bossy, less grown-up and less good-looking than those who were physically accelerated.

On the basis of these findings, it may be inferred that adult and peer attitudes toward the adolescent, as well as their treatment and acceptance of him, are related to his physical status. This means that the socio-psychological environment to which later-maturers are subjected–and consequently the social learning situations they encounter–may be significantly different from that of their early-maturing peers. As a consequence, according to the ratings summarized above, they acquire different patterns of overt social behavior. It seems reasonable to hypothesize that groups differing in physical status will also differ in more covert aspects of behavior and personality.

Indirect evidence relevant to this hypothesis comes from an investigation of the long-term consequences of physical acceleration or retardation during adolescence. Jones (1957) found that group differences in physique had practically disappeared by the time her early- and late-maturing subjects reached their early thirties. Nevertheless, young adults who had been physically retarded adolescents differed from those who had been accele-

---

Source: *Child Development*, 1957, *28*, 243–56.

rated in several important psychological characteristics. In general, it appeared that the adult subjects could be described much as they had been during adolescence. Thus, those who had been early-maturers scored higher on the good impression, socialization, dominance, self-control (low score on impulsivity), and responsibility scales of the California Personality Inventory, while those who had been slow in maturing scored higher on the flexibility scale. On the Edwards Personal Preference Schedule, early-maturers scored significantly higher on the dominance scale, while the late-maturing were high in succorance. Jones concludes that the early-maturing 'present a consistently favorable personality picture with regard to . . . important social variables'. Moreover, there was some evidence that these men had attained more stable vocational adjustments than those who had been late in maturing. These group differences in later adjustment suggest that the socio-psychological atmosphere in which the adolescent lives may have profound immediate and enduring effects on his personality structure as well as on his overt behavior.

The present study was designed to investigate the relationship between maturational status and certain important, covert aspects of personality during late adolescence. Personality structure was assessed by means of the Thematic Apperception Test (TAT), which seems to be the most appropriate and sensitive instrument for this purpose. More specifically, on the basis of the literature reviewed above and other general works on the psychology of adolescence (Farnham, 1951; Jones, 1943; Jones, 1949), we formulated and tested a series of propositions relating to differences between the physically retarded and the accelerated in self-conceptions, underlying motivations, and basic interpersonal attitudes. These variables were translated into TAT categories–needs (*n*), press (*p*), and descriptions (defined briefly in Table 1)–and the score of early- and late-maturers in each of these categories were compared. The propositions and the rationale underlying them, together with the TAT variables involved, follow.

1. In view of their obvious physical retardation, relatively unfavorable reputations and disadvantageous competitive position in many activities, the late-maturing boys are more likely to have feelings of inadequacy. Hence, more boys in this group than in the early-maturing group are likely to have negative self-conceptions (TAT category: *negative characteristics*).

2. The adolescent in our culture generally desires more independence and adult status. This may be the source of a major problem for the late-maturer, however, since he is often regarded and treated as a small boy by adults and peers and is not likely to be granted independence as early as physically accelerated boys. Therefore, it may be anticipated that more late- than early-maturers regard adults, particularly their parents, as dominating, forcing them to do things they don't want to or preventing them from doing things they want to do (high scores in *p Dominance*). Moreover, the parental treatment these boys experience and parental refusal to grant them

independent status may be interpreted as personal rejection. Hence, we predicted that more late-maturing boys would score high in *p Rejection*.

3. These feelings of being dominated and rejected may result in attitudes of rebellion against the family and in feelings of hostility. We therefore expected that more of the late-maturing group would reveal strong aggressive needs (high scores in *n Aggression*) and desires to escape from (*n Autonomy–leaving parents*), or to defy, the family (*n Autonomy–defying parents*).

4. On the basis of the data indicating that slow-maturers showed a great deal of social interest (although often of an immature kind), we hypothesized that more members of this, than of the early-maturing group would reveal strong interests in friendly, intimate interpersonal relationships (high scores in *n Affiliation*).

5. Assuming that, as Jones and Bayley (1950) suggest, the social initiative and attention-getting devices of the late-maturers are of a compensatory nature, we would expect this group to be basically dependent and to have strong needs for support from others. These should be manifest by higher scores in TAT *n Succorance* and *p Nurturance*. The latter may be considered a most indirect measure of dependence, a kind of wish-fulfilling view of the world as helpful and friendly.

6. The early-maturer, being regarded and treated as more adult, is more likely to become self-confident, and to acquire high status goals. For these reasons, we predicted that more of the physically accelerated would give evidence of high achievement goals (high scores in *n Achievement*) and concern with personal recognition (high scores in *n Recognition*).

7. Late-maturing boys in our culture probably face more problems of personal adjustment than do their early-maturing peers. As a result of this, they may become more aware of their problems, and, as the high degree of flexibility of young adults who had been retarded in maturing suggests, more insightful. Hence we predicted that they would be more willing and able than early-maturers to face their own feelings and emotions (low scores in the TAT variable *denial of feeling*).

In summary, we attempted to test seven propositions related to differences in the personalities of early- and late-maturing boys. It was hypothesized that more late-maturers would score high in variables relating to negative self-conceptions, dependence, aggression, affiliation, rebelliousness, and feelings of being dominated and rejected. More early-maturers, on the other hand, were expected to reveal strong achievement and recognition needs, feelings of personal success, and tendencies toward denial of feelings.

Procedure

The thirty-three seventeen-year-old male subjects of this investigation were members of the Adolescent Growth Study which included a normal sample of boys in an urban public school system (Jones, 1940). The subjects of the present investigation represented two contrasting groups, selected on the basis of their physical maturity status: sixteen of them had been among the most consistently accelerated throughout the adolescent period; the other seventeen had been among the most consistently retarded.[1] All of them took the Thematic Apperception Test, which provides the basic data of this study, at age seventeen.

The TAT consisted of eighteen pictures: nine from the Murray set which is now standard (cards 1, 5, 6, 7BM, 10, 11, 14, 15, 17); five pictures from the set generally used in 1938 when these data were collected (a man and woman seated on a park bench; a bearded old man writing in an open book; a thin, sullen, young man standing behind a well-dressed older man; a tea table and two chairs; an abstract drawing of two bearded men); and four designed especially for this investigation (the nave of a large church; a madonna and child; a dramatic view of mountains; a boy gazing at a cross which is wreathed in clouds).

The tests were administered individually. Each card was projected on a screen while the subject told a story which was recorded verbatim. Standard instructions were given for the Murray cards, and subjects were asked to describe the feelings elicited by the other four pictures. Most of the stories were brief, consisting of only one or two sentences.

As we noted earlier, each of the personality variables involved in the seven propositions was translated into a TAT scoring category. The scoring scheme involved counting the relevant needs, press, and description of the heroes of the stories, the assumption being that the storyteller has identified with the hero: the hero's needs are the same as the boy's; the press that impinge upon the hero are the ones that affect the boy telling the story. A total of twenty needs, press, and descriptive categories, each defined as specifically as possible, was developed in the analysis of the protocols. A score for each subject for each TAT category was derived by counting the number of stories in which it appeared. A list of the categories used, together with brief descriptions of them, is found in Table 1.

To test the reliability of this analysis, one of the authors (PM) and another psychologist (Dr Virginia B. Ware) independently scored fifteen complete protocols (300 stories). The percentage of interrater agreement was 90, computed by the usual formula (number of agreements divided by number of agreements plus number of disagreements).

In order to eliminate bias, the scoring used in the present study was done

---

[1] The present sample includes 27 of Jones and Bayley's 32 subjects (the 16 most consistently retarded and 16 most consistently accelerated boys in the study). The other five boys had not taken the TAT at age 17. The six subjects who were in the present study but not in Jones and Bayley's study are the three 'runners-up' from each end of the physical maturity distribution, i.e. the three who were closest to the 16 most accelerated cases and the three cases next to the 16 most retarded.

'blind', that is, independently of knowledge of the subject's maturational status.

Results

Frequency distributions of the scores of all subjects were made for all the TAT variables. Each distribution was then dichotomized at the point which most nearly enabled the placing of half of the thirty-three subjects above, and half of them below, the dividing point. Subjects having scores above this point were considered high in this particular variable; those with scores below that point were considered low in this variable. Chi square tests were used to test the seven propositions, i.e. to ascertain whether or not high scores in certain TAT variables were in fact more characteristic of one group (late- or early-maturers) than of the other.

Table 1 lists the TAT variables, the number of late- and early-maturers with high scores in the variable, the chi square value obtained and the level of significance. It should be noted that the hypotheses tested were one-sided hypotheses, while the chi square value is in terms of a two-sided hypothesis. When chi square has only one degree of freedom, the square root of chi square has a distribution which is the right hand half of a normal distribution. In order to test a one-sided hypothesis, the chi square test must be converted into the equivalent value in terms of a unit normal deviate (Fisher 1938). The levels of significance reported in Table 1 were evaluated in these terms.

Table 1 shows that, as had been predicted, more late-maturing than early-maturing boys revealed feelings of inadequacy and negative self-concepts, i.e. scored high in the TAT variable *negative characteristics*. Hence proposition 1 was confirmed. This finding is consistent with the frequently made clinical observation that retardation in physical maturation may be an important source of personal maladjustments and attitudes of inferiority.

Proposition 2 stated that more late-maturers regarded their parents as highly dominating and rejecting. The evidence summarized in Table 1 substantially supported this proposition. While the difference was not statistically significant, more late- than early-maturers scored high in *p Dominance by parents* (total). There was a marked difference between the groups in the variable which involves parental domination by forcing the child to do something he does not want to do (*p Dominance by parents, forcing*). However, examination of the data with respect to the variable *p Dominance by parents (prevention)* makes it necessary to reject that part of the proposition which maintains that late-maturers are more likely to view their parents as highly restrictive in their activities.

That aspect of proposition 2 which deals with feelings of rejection was confirmed by our data. Compared with the early-maturing group, a significantly greater proportion of the late-maturers told stories in which the

Table 1 *Number of early- and late-maturers scoring high in TAT variables*

TAT variable	*Definition of variable*	*High early-maturers*	*High late-maturers*	*Chi square value*	*p*
Proposition 1					
Negative Characteristics	H* is described in negative terms (e.g. imbecile, weakling, fanatic)	5	13	6.80	<0.01
Proposition 2					
p Dominance 1	H forced by parents to do something he doesn't want to	4	8	1.73	0.09
p Dominance 2	If prevented by parents from doing something he wants to	6	8	0.31	>0.30
p Dominance 3	Total instances of H's being forced by parents to do something and/or prevented from doing something	7	11	1.46	0.11
p Rejection	H rejected, scorned, or disapproved of by parents or authorities	5	11	3.69	0.03
Proposition 3					
n Agression 1	H is aggressive in physical, asocial way	8	3	3.88	0.02
n Agression 2	H is mad at someone, argues	7	4	1.52	0.10
n Agression 3	Total of all H's aggressive actions	11	8	1.26	0.10
n Autonomy 1	H leaves home	7	10	0.75	0.20
n Autonomy 2	H disobeys or defies parents	7	11	1.46	0.11
n Autonomy 3	Total of instances in which hero leaves and/or defies his parents	3	9	4.16	0.02
Proposition 4					
n Affiliation 1	H establishes good relations with his parents	8	8	0.00	>0.50
n Affiliation 2	H falls in love, has a romance, marries	9	14	2.66	0.05
n Affiliation 3	Total instances in which H establishes and/or maintains friendly relations	8	12	1.46	0.11
Proposition 5					
n Succorance	H feels helpless, seeks aid or sympathy	7	12	2.43	0.06
p Nurturance 1	H is helped, encouraged, or given something by parents	5	8	0.93	0.18
p Nurturance 2	H is helped, encouraged, or given something by someone else (not parents)	8	14	3.88	0.02

**H stands for hero.*

Proposition 6					
n Achievement	H attempts to attain a high goal or to do something creditable	9	10	0.02	>0.50
n Recognition	H seeks fame and/or high prestige status	9	8	0.28	>0.30
Proposition 7					
Denial of Feeling	S states that picture elicits no thoughts or feelings	9	5	2.43	0.06

hero was rejected by parents or authority figures. These feelings of rejection may stem from different sources. In some cases, the parents' behavior may make it clear that they are disappointed in their physically retarded son whom they regard as immature. The boy, perceiving this attitude, may interpret it as rejection. In other cases, parental reluctance to allow the late-maturing boy to establish his independence may lead to considerable tension in the family and the boy's feelings of rejection may simply reflect the ongoing parent-child conflict.

It is possible that earlier in their teens, soon after the physical changes of adolescence became apparent, many of the early-maturing boys also experienced conflicts with their parents, arising from difficulties in establishing their independence or in handling emerging heterosexual interests. At that time they too may have felt dominated or rejected. However, by the age of seventeen, when these data were collected, these boys were ordinarily treated as adults and granted more freedom. Hence, they were more likely to have resolved many of their conflicts with their parents and to feel accepted and independent.

The hypothesis (part of proposition 3) that more late-maturers would be highly aggressive was rejected on the basis of the evidence given in Table 1. In fact, the differences between the two groups on all the TAT aggression variables were in the opposite direction from the prediction. High scores in the variables relating to aggression of the most overt and violent type were significantly more frequent among the early-maturers, and more members of this group also scored high in measures of milder (verbal) aggression and of total aggression. While late-maturers may experience more problems of adjustment and greater frustrations than their early-maturing peers, they apparently do not manifest greater aggressive motivation. It may be that their own feelings of inadequacy or fears of retaliation and punishment for aggression inhibit their expression of hostile feelings, even in fantasy. On the other hand, the early-maturers who feel more secure personally, and recognize their own relatively advantageous physical and social status, may feel freer to express their aggressive needs. Since aggression is a culturally stereotyped masculine trait, it seems possible that the physically accelerated, being accepted as mature and identifying readily with adult males, are more likely to acquire this characteristic. In any case, the finding that early-maturers express higher aggressive motivation during late adolescence seems consistent with Jones' finding that, as young adults, they score high on the dominance scale of the Edwards Personal Preference test (Jones, 1957). Perhaps the relatively strong aggressive motivation of the early-maturer, or the mature sex-role identification it may imply, serves as a basis for the development of later qualities of leadership and persuasiveness (Jones, 1959).

As Table 1 indicates, the other aspect of proposition 3 was confirmed: a significantly greater proportion of late- than of early-maturers displayed strong motivations to escape from, or defy, their parents. These may be essentially aggressive reactions, stemming from feelings of parental domination and rejection, or they may reflect the late-maturers' awareness of their strife with their parents whom they perceive as blocking their drives for independence. These strong needs for escape and defiance may also be considered evidence of a generally immature way of handling parent-child conflicts. Perhaps, by the age of seventeen, the early-maturers have already resolved many of their conflicts with their families and/or have learned to handle these in less rebellious and in more direct and mature ways.

Proposition 4 stated that, compared with their early-maturing peers, more late-maturers would manifest strong needs for establishing close social contacts with others. While there was some confirmatory evidence, the results were not clear-cut. When all affiliative needs were considered together (score for *n Affiliation–total*), the group differences were in the predicted direction, but not statistically significant. Examination of the protocols revealed that almost all instances of affiliation concerned either parents or the opposite sex; there were very few stories involving close, friendly associations between like-sexed peers. The two major types of affiliation were scored separately. As Table 1 shows, late-maturers did not differ from early-maturers with respect to need for affiliation with parents, but a significantly greater proportion of the former group displayed strong motivation for heterosexual affiliation.

In view of the late-maturers' strong feelings of inadequacy and dependent needs (see below), it is surprising that a greater proportion of this group did not exhibit strong needs to establish and maintain close bonds with their parents. This may be due to the late-maturers' more intense conflicts with their parents at this age (seventeen years), their fears of being rejected and dominated by them, and their generally defiant attitudes which prevent them from admitting, even in fantasy, their strong underlying needs to form close contacts with them.

The significant difference between the groups in *n Affiliation* (*love, romance, marriage*) is subject to several possible interpretations. For one thing, this category may refer to general needs to establish close relations with others (with peers or adults other than parents) and not merely to desire for contact with the opposite sex. The set of stimulus cards may not have been adequate to elicit responses indicative of more general affiliative needs; hence, these were expressed through responses in the heterosexual affiliation category. If this is true, proposition 4 was confirmed, and the late-maturers' high scores in this variable indicate their greater general interest in establishing and maintaining friendly relationships.

It is also possible that the late-maturers' strong affiliative needs are actually directed only toward members of the opposite sex, i.e. that *n Affiliation* (*love, romance, marriage*) measures specifically heterosexual interest. Assuming that this is true, there is another plausible explanation for the discovered difference. As we saw earlier, the late-maturer may be afraid to admit that he desires close association with his parents. He may also feel that his immaturity and poor reputational status prevent him from establishing successful social relationships with like-sexed peers. Hence, he may 'displace' his affiliative needs to members of the opposite sex, who, in his fantasies, may seem more responsive.

A third possible explanation of the difference is based on Jones and Bayley's findings that the late-maturers show less overt interest in girls and are regarded as less good-looking. From these data, it may be inferred that the physically retarded probably do not have successful and rewarding experiences with girls. Hence their heightened need for affiliation with the opposite sex, expressed in the TAT, may reflect their attempts to satisfy in fantasy needs which they cannot satisfy adequately in reality.

The data were generally supportive of proposition 5 which stated that late-maturers are likely to have strong underlying dependent needs. A higher proportion of this group than of their early-maturing peers scored high in *n Succorance,* the difference between the two groups approaching statistical significance (p=0.06). Furthermore, high scores in the category involving receiving help and support from others (not including parents) (*p Nurturance–non-parents*)–an indirect measure of dependent needs–were significantly more characteristic of the physically retarded than of the physically accelerated. In view of the late-maturers' attitudes toward their parents, discussed above, it is not surprising to find that perceptions of parents as kindly and supportive (high scores in *p Nurturance–parents*) were not significantly more common in this group than in the early-maturing group.

On the basis of the data involving the TAT variables *n Achievement* and *n Recognition,* we rejected proposition 6 which stated that more early-maturers would be self-confident and have high needs for achievement and personal recognition. In our culture there is strong pressure to develop needs for achievement and personal recognition, and, according to our results, these needs and feelings may become intense regardless of–or perhaps in spite of–the child's maturational status, feelings of personal adequacy, dependency, and adjustment to parents.

Two interesting incidental findings from the TAT data seem to be consistent with the proposition that more early- than late-maturers are likely to be self-confident. Seven boys in this sample of thirty-three adolescents told stories in which the hero was helpful or kind to someone else (*n Nurturance*). Of this group, six were early-maturers, while only one was a late-maturer (x^2=2.09, p=0.07). In so far as *n Nurturance* may be a measure of the storyteller's own feelings that he can accept an active, mature role, more of the accelerated group feel self-assured with respect to having attained mature status.

The other incidental finding which seems to support proposition 6 is based on responses only to card 1 of the Murray series which depicts a young boy contemplating a violin which rests on a table in front of him. Eight of the subjects spoke of the boy (the hero) as a prodigy or a genius. Of these, seven were early-maturers; only one was physically retarded (x^2=5.25, p=0.01). If the attribution of this prestige status and accomplishment to the hero reflects the subject's own feeling that he has been an achiever, it follows that more of the physically accelerated have positive self-concepts. In view of the small number of cases involved, both of these findings must be considered tentative, but they do offer some evidence in support of proposition 6.

Proposition 7, which stated that relatively few of the physically retarded boys are unwilling or unable to face their own feelings and emotions, received some support from the TAT data summarized in Table 1. A smaller proportion of the members of this group than of the physically accelerated group specifically denied that the pictures evoked any feelings or emotions (e.g. 'It doesn't make me think of anything'). While this variable may not adequately measure *denial of feeling* as the major defense mechanism, this result seems to indicate that late-maturers are more sensitive to their own feelings and more ready to admit and face them openly. Since these qualities are basic to the development of psychological insight, it may be inferred that late-maturers, as a group, are more likely to become insightful individuals.

Discussion

The results of the study support the general hypothesis that, in our culture, the boy whose physical development is retarded is exposed to a socio-psychological environment which may have adverse effects on his personality development. Apparently, being in a disadvantageous competitive position in athletic activities, as well as being regarded and treated as immature by others, may lead to negative self-conceptions, heightened feelings of rejection by others, prolonged dependent needs, and rebellious attitudes toward parents. Hence, the physically retarded boy is more likely than his early-maturing peer to be personally and socially maladjusted during late adolescence. Moreover, some of his attitudes are likely to interfere with the process of identification with his parents, which is generally based on perceptions of them as warm and accepting (Payne and Mussen, 1956). This, in turn, may inhibit or delay the acquisition of mature characteristics and attitudes which are ordinarily established through identification with parents. Fortunately for the late-maturers' subsequent adjustments, they seem more willing and able to face their feelings and emotions. This may be a result of their awareness of others' attitudes toward their immaturity or their feelings of personal inadequacy and dependency.

The physically accelerated boys, on the other hand, are likely to experience environmental circumstances which are much more conducive to good psychological adjustment. Hence, their psychological picture, as reflected in their TAT stories, is much more favorable. By the time they were seventeen, relatively few early-maturers harbored strong feelings of inadequacy, perceived themselves as rejected or dominated by parents or authorities, or felt rebellious toward their families. As a group, they appeared to have acquired more self-confidence and had probably made stronger identifications with mature adults. Hence, they perceived themselves as more mature individuals, less dependent and in need of help, and more capable of playing an adult male role in interpersonal relationships.

These findings assume additional, probably greater, importance when they are considered in the light of Jones' findings on the early adult (age thirty-three) adjustments of boys who had been retarded or accelerated in physical maturing (Jones, 1957). It should be recalled that by this age physical differences between the two groups had practically disappeared. Certain important psychological differences were noted, however, and these were consistent with the differences at age seventeen, reported in the present study. For example, the responses of the early-maturing group to two paper-and-pencil tests, revealed that, as young adults, they were more dominant, more able to make a good impression and more likely to be turned to for advice and reassurance; more self-controlled; and more willing and able to carry social responsibility. In short, they present a general picture of psychological maturity. Moreover, more of the early-maturers seemed to have made successful vocational adjustments. In contrast to this, when the late-maturers became adults, they tended to be highly dependent individuals who could be described, on the basis of their test responses, as tending to be rebellious, touchy, impulsive, self-indulgent, and insightful. Most of these characteristics are indicative of poor adjustment and psychological immaturity. Fewer members of this group had made good vocational adjustments.

The striking correspondence between the two descriptions of the groups, derived from different kinds of tests and collected at widely separated periods of time, lends further support to Jones' (1957) conclusion that 'the adolescent handicaps and advantages associated with late- or early-maturing appear to carry over into adulthood to some extent'. It seems clear that many attitudes of adolescent personality (patterns of motivation, self-conceptions, and attitudes toward others) characteristic of late- and early-maturing boys are relatively stable and durable rather than situational and transitory. This may be attributable to the fact that in our culture adolescence is generally a critical and difficult period of adjustment. Within a relatively brief interval of time, the child must work out numerous complex and vitally important personal problems—e.g. adaptation to his changed biological and social status, establishment of independence, vocational adjustment. In dealing with these problems, he may acquire new behaviors and personality attributes which have broad ramifications, not only on his current adjustment, but also on his subsequent development. If the adolescent can cope with his problems without too much inner stress and turmoil, his self-esteem, feelings of adequacy, and consequently his subsequent adjustment, are likely to be enhanced. On the other hand, if his problems induce great tension and anxiety, he is likely to feel frustrated and inadequate, and, if these feelings are maintained, to adjust less satisfactorily as an adult.

Obviously, the adolescent's success or failure, as well as ease or tension, in handling his problems will be determined to a large degree by the socio-psychological forces to which he is subjected during this time, and these, as we have seen, may be significantly related to his rate of maturation. Thus, physical status during adolescence—mediated through the socio-psychological environment—may exert profound and lasting influences on personality. For this reason, many aspects of the adult's behavior and personality seem consistent with his adolescent adjustments, attitudes and motivations.

In so far as our results permit generalization, they suggest that some important aspects of motivation, such as needs for achievement and personal recognition, are not significantly affected by maturational status. It may be that among subjects whose achievements are strongly encouraged and rewarded from very early childhood, the need to achieve becomes powerful and resistant to change even in the face of feelings of helplessness and inadequacy. The latter may inhibit the achievement-oriented overt behavior of some late-maturers, but the underlying motivation to achieve seems as strong in this group as it is among the physically accelerated.

In conclusion, it should be noted that, although rate of maturing and associated factors may affect personality development, the relationship between physical status and psychological characteristics is by no means simple. A vast number of complex, interacting factors, including rate of maturation, determines each adolescent's unique personality structure. Hence, in any specific instance, the *group* findings of the present study may not be directly applicable, for other physical, psychological, or social factors may attenuate the effects of late- or early-maturing. For example, an adolescent boy who is fundamentally secure and has warm, accepting parents and generally rewarding social relationships may not develop strong feelings of inadequacy even if he matures slowly. Analogously, the early-maturing boy who has deep feelings of insecurity, for whatever reasons, will probably not gain self-confidence simply because he matures early. In summary, in understanding any individual case, generalizations based on the data of the present study must be particularized in the light of the individual's past history and present circumstances.

References

FARNHAM, M. L. (1951) *The adolescent.* New York: Harper.

FISHER, R. A. (1938) *Statistical methods for research workers.* (7th Ed.) Edinburgh: Oliver & Boyd.

JONES, H.E. (1940) Observational methods in the study of individual development; *J. Consult. Psychol., 4,* 234.

JONES, H.E. (1943) *Development in adolescence* New York: Appleton-Century.

JONES, H. E. (1949) 'Adolescence in our society'. In *The family in a democratic society, Anniversary papers of The Community Service Society of New York.* New York: Columbia University Press.

JONES, M. C. (1957) 'The later careers of boys who were early- or late-maturing', *Child Developm., 28,* 113–28.

JONES, M. C. (1959) 'A study of socialization at the high school level', *J. genet. Psychol., 93,* 87–111.

JONES, M. C., and BAYLEY, N., (1950) 'Physical maturing among boys as related to behavior; *J. Educ. Psychol., 41,* 129–48.

PAYNE, D. E., and MUSSEN, P.H. (1956) 'Parent-child relations and father identification among adolescent boys', *J. abnorm. soc. Psychol., 52,* 358–62.

STONE, C. P., and BAKER, R. G., (1939) 'The attitudes and interests of premenarcheal and postmenarcheal girls', *J. genet. Psychol., 54,* 27–71.

TRYON, M. (1939) 'Evaluation of adolescent personality by adolescents', *Monogr. Soc. Res. Child Developm., 4,* No. 4.

17 On equivalence

R. R. Olver and J. Rigney Hornsby

That children and adults group discriminately different things and treat them as 'the same' or 'alike' is hardly debatable. And indeed, were they not naturally prone to do so, the diversity of the environment would soon overwhelm them. Such equivalence-making is in large measure a learned achievement; it may be expected to change with growth and development in a manner consistent with more general changes in cognitive development. Elsewhere we have made a good deal of how the very young child represents and 'knows' by doing, how there is added to this primitive mode a capacity for knowing by depicting in images, and finally how the growing child achieves the ability to give an account of his world in the powerful medium of language. This course of growth by which finally all three techniques of knowing come into force–enactive, ikonic, and symbolic representation–is reflected in the changing ways that children have for imposing equivalence on the things of their world.

Enactive, ikonic, and symbolic representation might each, for example, be expected to emphasize different features of the environment as bases for establishing equivalence. Under enactive representation, things should be seen as alike on the basis of a common role in some action. Equivalence with ikonic representation might more likely be accomplished by grouping items according to perceptual kinship or likeness. With the achievement of symbolic representation, equivalence might well be expected to be governed by such grammatical principles as synonymy, superordination, or syntatic substitutability. It is with such a course of growth that this article deals. The transitions with which we shall be concerned are interesting in their own right as a documentary on development. More interesting still is the picture of the underlying form of organization in thought that is revealed.

For example, it is not only the 'semantics' of equivalence that changes with growth–the features of the environment used as the basis of equivalence–but the 'syntax' of equivalence formation as well. Might equivalence groupings under enactive representation demonstrate the sequential properties of action sequences? Might groupings governed by ikonic representation, reflecting the domination of the perceptually vivid be based on a kind of conjunctive joining of attributes? Does the development of symbolic equivalence groupings take on the form of conventional categorization and hierarchical organization?

The studies reported here trace the development of equivalence from age

---

Source: in BRUNER, J. S. *et al.* (eds.) (1966) *Studies in Cognitive Growth.* New York: John Wiley, 68–85.

six to age nineteen, from the first year of regular school work to the first year in college. The materials we have used permit the child to demonstrate both the basis on which he renders things equivalent and the syntax or structure of the groups he forms.

Experiment 1: Equivalence formation with verbal materials

We gave children from age six to nineteen the task of telling us how different items are alike. We presented the words *banana* and *peach*, each typed on a small white card and spoken aloud as well, and asked the child, 'How are banana and peach alike?' We then added *potato* to the list, first asking, 'How is potato different from banana and peach?' and then, 'how are banana, peach, and potato all alike?' Next we added *meat*, asking, 'How is meat different from banana, peach, and potato?' and then, 'How are banana, peach, potato, and meat all alike?' This procedure was continued until the array consisted of: *banana, peach, potato, meat, milk, water, air, germs.* At the end of the array we included an item about which we asked only how it differed from the preceding items; for example, *stones* was presented as the final item in the banana-peach list. We presented a second array of items in the same manner: *bell, horn, telephone, radio, newspaper, book, painting, education,* and as the contrast item, *confusion.*

Note that the arrays are made up of successively more distant items, but, though the items became increasingly more diverse, they share a common characteristic. The items in the banana-peach array, for example, are all ingestible; those in the bell-horn array all carry messages, and so forth. As words are added, the task gets more difficult, and so it was intended, for we were interested in pushing our subjects to their limits. The two lists were constructed impressionistically. Could we have found a more systematic method–we tried several–it would have been better. The present lists do at least evoke responses that serve our purpose.

Sixty children formed the subjects in this investigation. The children in the youngest grades–from age six to fourteen–were enrolled in surburban public schools near Boston; the sixteen-year-olds came from a nearby suburban high school; and the oldest group (the college freshmen) were students of Harvard and Radcliffe Colleges. There were five boys and five girls in each age group, and their mean ages and IQ scores are shown in Table 1.

Children of different ages go about such a task in rather different ways. One six-year-old said that banana, peach, potato, and meat were alike because (a banana is yellow, and a peach is red and yellow, or sometimes red and sometimes just yellow, and a potato is light flesh, and meat is brown'. Another responded, 'Meat and potato are most of the time together, peach and banana can be for dessert, and the banana is close to the kind of vegetable'. Older children, for example, the sixteen-year-olds grouped the

Table 1 *Subjects in Experiment 1–ages, IQs, number of children*

Grade	*Mean Age*	*Mean IQ*	*Number*
1	6 years 3 months	*	10
4	9 years 6 months	122	10
6	11 years 7 months	115	10
8	13 years 5 months	122	10
10	15 years 11 months	122	10
Freshmen	18 years 7 months	*	10

* IQ data not available

same items as 'They're all something to eat', or 'They're all food'. The six-year-old, consistent with his reliance on ikonic representation, seemed caught by the way things look–their color or where they are found together. The sixteen-year-old spoke of the function of the items–what they are used for. Not only did the bases for grouping differ, but the syntax as well. While the older child used a common characteristic to form a class, some property shared by all members, the six-year-old linked each item only to the next one, or stated a characteristic separately for each item in the group.

Consider more closely now on what bases the items are judged to be 'the same'. Since they all vary from one another in a number of different dimensions, a variety of different characteristics could be used as the basis of equivalence. Five main modes can be distinguished: *perceptible, functional, affective, nominal,* and *fiat equivalence.* For each mode we can set forth a typical language frame for characterizing the basis of equivalence.

1. *Perceptible*: The child may render the items equivalent on the basis of immediate phenomenal qualities such as colour, size, shape, or on the basis of position in time or space.

Perceptible Intrinsic	They are ——. (X:adjective: '. . . both yellow'.) They have ——. (X:noun: '. . . writing on them'.)
Perceptible Extrinsic	They are (preposition) ——. (X:position in time or space: '. . . all in a house'.)

2. *Functional*: The child may base equivalence on the use or function of the *items, considering either what they do or what can be done to them.*

Functional Intrinsic	They ——. (X:verb: '. . . make noise'.)
Functional Extrinsic	You —— them. (X:verb: '. . . can turn them on'.)

3. *Affective*: The child may render the items equivalent on the basis of the emotion they arouse or of his evaluation of them.

Affective You——them. (X:value or internal state: '. . . like them both'.)
They are——. (X:adjective indicating value: '. . . very important'.)

4. *Nominal*: The child may group the items by giving a name that exists ready-made in the language.

Nominal They are (or are not)——. (X:noun: '. . . both fruit'.)

5. *Fiat Equivalence*: The child may merely state that the items are alike or are the same without giving any further information as to the basis of his grouping, even when he is prodded.

Fiat Equivalence 'A' is (or is not)——'B'. (X:like, similar to, the same as, and so forth: 'They are the same thing, really'.)

Six-year-olds do, indeed, group more often according to perceptible properties than do older children. Their protocols are laced with the colors, sizes, shapes, and places of things. More than a quarter of their groupings are of the perceptible intrinsic type; no older group forms even half that many on this basis. From age six on there is a steady increase in functionally based equivalence–from 49 per cent of all responses at age six to 73 per cent at age nineteen (Figure 1). Initially the turn to functional

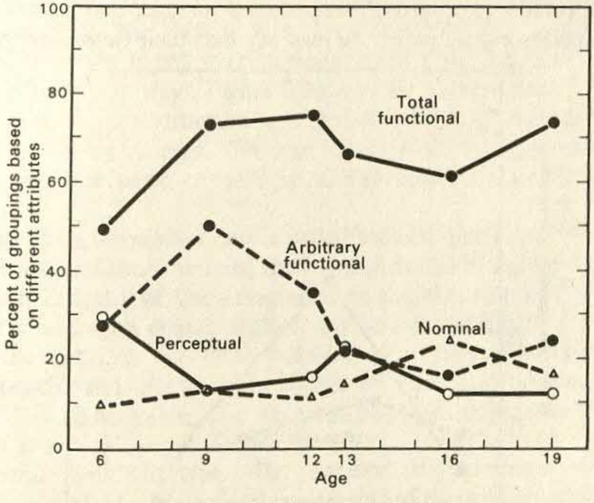


Fig. 1 Percent of groupings based on different attributes

attributes as the basis of equivalence is extrinsic as we have defined it, and often it is seemingly arbitrary–the child speaks of what *he* can do *to* the objects: turn them on, roll them up into a ball, and so forth. At age nine half of the children's groupings are formed in this way. It is as if the child were taking himself as reference point in order to create a common basis for grouping an apparently diverse array.

What emerges at this point is a bit surprising. One is at first taken aback by the central role of 'functionalism' as an arm or at least an accompaniment to intellectual growth. It has about it some of the properties of enactive representation–definition by an action taken. But a closer examination suggests that this new functionalism provides a means for the child to get free of responding to the more immediate, surfacy aspects of the things around him. Indeed, functionalism in its egocentric form is what permits the child to distinguish between objects and actions taken toward them. Yet a closer look at the syntactic aspect of grouping as it expresses itself at this time in the child's life–and we shall have a closer look presently–also makes it plain that functionalism is highly associated with more mature forms of grouping. It may well be that as the child breaks away from the perceptual domination of vivid things he must fall back on a more practical way of dealing with the environment–through action, or at least vicarious action. The common uses of things are pitted against their divergent appearances, and the conflict promotes growth. Whether functionalism is made possible by the child's ability to use new grouping principles, or the other way round, must remain a moot point.

Consider further this increase in grouping according to use, even arbitrary use. The nine-year-old, for all his functionalism, forms groupings that are neither particularly appropriate nor realistic for adaptive action. Without intending a philosophical point, we may say that their functional groups are

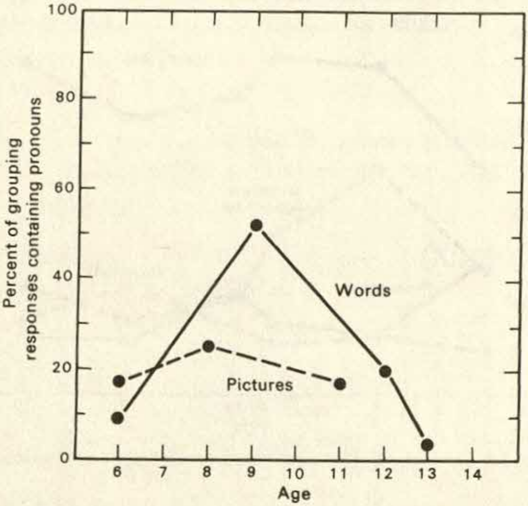
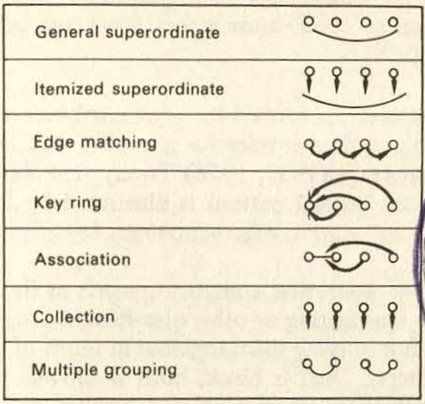


Fig. 2 Percent of grouping responses containing pronouns You and I–verbal and picture array



STATE INSTITUTE OF EDU
Govt of West Ben

Fig. 3 Grouping syntax diagrams

often arbitrary and that they ignore the conventional or sensible uses to which objects are put. The child at this age ignores the possible reciprocal relations between himself and the objects he is sorting and, instead, imposes functions upon them. There is, for example, a sharp increase at age nine in the use of the pronouns 'you' and 'I' in answering the questions we pose. And the use of 'you' and 'I' in framing responses declines beyond age nine as the child turns from extrinsic to intrinsic functional properties as the basis of grouping. The shift is from responses such as, ' can crinkle up a newspaper and then it will make a noise like the bell and horn' to 'They all tell ideas in their own way'. Figure 2 depicts the relevant data.

Consider now the structure or 'syntax' of the groupings formed by children of different ages. We can distinguish three general grouping structures: *superordinate, complexive,* and *thematic* (Figure 3).

1. *Superordinate groupings* are constructed on the basis of a common feature or features characterizing the items included in a group or class. This is the classic category of Venn diagrams and the like. Any array of items has a number of common characteristics, any one or combination of which can serve as the criterion for their inclusion in a group. Thus, for example, banana, peach, and potato can be placed in a superordinate grouping because 'They all have skins', or 'They are all food', or because 'They all can be bought at a store', and so forth.

General superordinate. The general superordinate construction consists of stating a common characteristic of the items in the group. For example, bell and horn are 'both things that make noise'.

Itemized superordinate. Itemization may be added to superordinate grouping such that, while the items have a generalized property that joins them, the basis on which each item qualifies is explicitly stated. For example, 'Bell makes noise, horn makes noise too, bell says ding-dong, horn says doo-doo'.

2. *Complexive structures* are formed by using attributes of an array so as to form local rather than universal rules for grouping and, in this sense, these are closer to Wittgenstein's (1953, 1958) 'family resemblances' than to the classic category.[1] This general pattern is illustrated by five maneuvers for forming complexes: *collections, edge matchings, key rings, associations,* and *multiple groupings.*

Collections. The collection complex consists in finding complementary or somehow contrasting or otherwise related properties that all the things have, but not in tying them together in terms of attributes that are shared. For example, 'Bell is black, horn is brown, telephone is blue, radio is red'. Or, 'Newspaper you can read, book you can read, telephone you get messages over, radio you get messages over, and a horn you can blow'.

Edge matchings. The edge-matching complex consists in forming associative links between neighboring items. A chain of items is formed by tying the items together in linked pairs. For example, 'Banana and peach are both yellow, peach and potato are round, potato and meat are served together, meat and milk both come from cows'. There is no consistency in the attribute or characteristic by which one link of the chain is joined with the one that precedes and the one that follows.

Key rings. The key-ring complex consists in taking an item and linking all the others to it by choosing attributes that form relations between the central item and each of the others. For example, 'Painting–well, one thing is a newspaper has got some painting in it, a book has got some black printing, a radio and a telephone have painting on them and a horn–well, there's a little painting on it, and a bell is also the color of paints'. Or, 'Germs are in banana, peach, potato, meat, milk, water, and air'.

Associations. In the association complex the child links two items and then used the bond between these items as a nucleus for the addition of other items. For example, 'Bell and horn are music things, when you dial a telephone it's music a little'. Or, 'Bell, horn, telephone, and radio make noises, if you fold back a newspaper, then it will crackle and make a noise'.

Multiple groupings. The multiple-grouping complex consists of the formation of several subgroupings. For example, 'A telephone is like a radio–I know that. A horn and a bell both make sounds, but I don't

---

[1] Wittgenstein is quite right in pointing out that such complexes as we describe have a useful status in scientific as in everyday thinking. We may at times in our discussion appear to denigrate complexive thinking, but that is not our intention. Rather, our concern is with the child's capacity to go beyond such grouping–not simply with whether he abandons or foreswears complexive grouping.

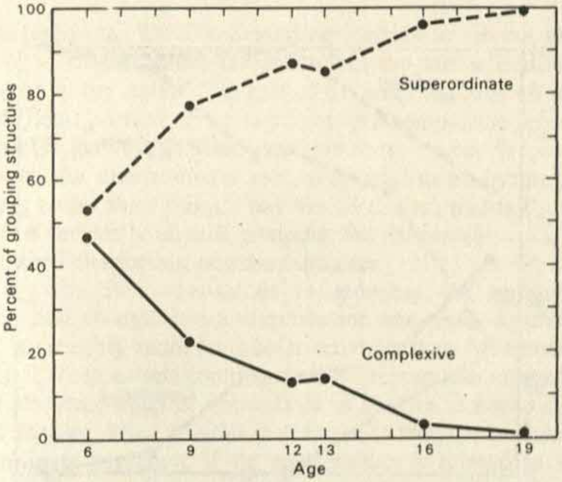


Fig. 4 Percent of grouping structures of two types

know about a newspaper'. The list is thus segmented into several groups, and the gap between them is not bridged.

3. *Thematic groupings* are formed on the basis of how the items fit in a sentence or a story or a thema. The construction of thematic groupings, in fact, most often depends on a sentence for tying items together. The sentence carries a story or thematic line: 'The little boy was eating a banana on the way to the store to buy some peaches and potatoes'.

We can turn now to the patterns of growth observed in the children. To begin with, there is a massive change that takes place between the sixth and nineteenth year (Figure 4). At six, half of the groupings made by the children are complexive, half superordinate. By nine, the balance shifts to three-quarters superordinated. And by ninteen, the complexive grouping has virtually disappeared, at least among these subjects and in this culture.

Not only do the younger children use complexive groupings more often, they also fall back on them more readily when the going gets hard. Recall that the items in an array diverged increasingly from the opening pair as successive ones are added to the list. The degree of divergence is an inexact matter, but at the very least it can be said that for virtually anybody, child and adult alike, it is more difficult to form a grouping for the first four or five items in a list than for the first pair. We took advantage of this fact to compare the performance of our subjects on the first pair, trio, and quadrad of items (easier items) with performance on five, six, seven, and eight items. The results tell their own story–brave starts and weak finishes by the younger subjects, until finally the task is within reach of the older ones (Figure 5). And when the younger children fall back on complexes with the longer lists, they tend to use the less demanding forms such as collections

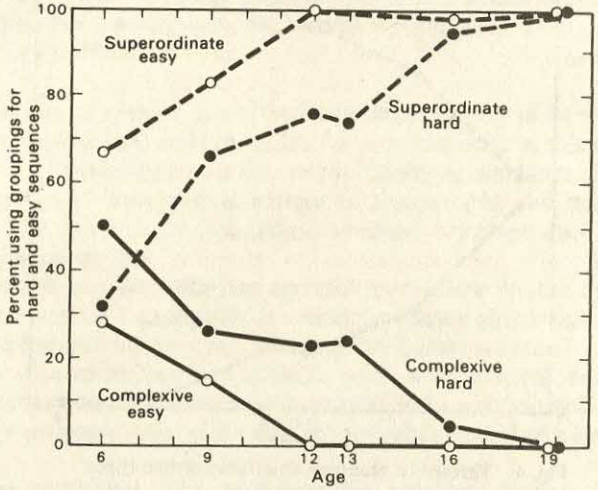


Fig. 5 Percent of groupings in easy and hard sequences that are complexive and superordinate in structure

and edge-matchings, whereas the older ones use associations and key rings more often.

There is a striking relation between the syntax and the semantics of the groupings the children made. If the attributes used in grouping are perceptible, then the grouping is likely to be complexive. If attributes are functional, on the other hand, the chances are greater that they will be grouped superordinately.

Our excursion into the domain of grouping, using verbal materials as stimuli, leads us to several conclusions. Perhaps the most obvious is that in tasks such as we presented our subjects, growth is reflected by a steady movement from grouping complexively to grouping by superordination. This progress is achieved at the same time that the child stops paying sole attention to the perceptible properties of the objects to be grasped and begins attending more to their functional significance.

We cannot help saying a word of caution. It is evident that the tasks we used were ones that the children were increasingly likely to see as they grew older as 'academic', to be dealt with 'intellectually', as one deals with such tasks. To a degree, their competence to deal with tasks in this way–consistently, logically, and so on–depends upon their being in the culture, whether in school or in social settings where we demanded of older children that they put aside childish ways of thinking. There is little question that our older subjects were also able to be more complexive, though they did not sense this experiment as an appropriate occasion for it.

We have observed both the 'overshoot' of complexive thinking in the younger children and the corresponding 'overshoot' of the use of superordinates in somewhat older children. Six-year-olds who attempt to use superordinates in the 'easier' first half of the array fall back on complexes in the more difficult portion of the task. Yet by a comparable token, nine- and twelve-year-olds get lost in what we have come to call 'hyperordination'. Trying to use the superordinate rule, they mire down by grouping highly diverse items under the rubric, 'They are all things', and fail to achieve the benefit of the simplicity of such grouping. For 'hyperordination' is knowing the form of the behavior but not the substance.

Obviously, the two approaches to grouping are required in adult functioning, and though in our data we see one replacing the other, the placement is probably more for public activities than for those done more subjectively. The loose-knit complex, as Wittgenstein and others have noted, is a vehicle for searching out possibilities of kinship. It is also the vehicle of poetry and fantasy. What it lacks in tidiness, it recovers in richness. So too the superordinate category: if its applicability is limited to well-formed problems, at least it is capable of precision and a workable exclusiveness.

Before drawing conclusions, we should examine a variant of the present experiment, lest we be victim of results imposed by our methods. We turn now to a related experiment much like the first, save in two respects: the first experiment used verbal materials, and it forced them in a fixed order on the children. To some extent, these procedures constrained the nature of the response made.

In the second experiment, the stimuli were pictorial, and the subjects could deal with them in their own order.

Experiment 2: Equivalence formation with pictures

We presented children from age six to eleven with an array of forty-two water-color drawings. Ninety boys, thirty each from grades 1, 3, 6, of a suburban school near Boston served as subjects, and each was individually tested (mean ages and IQs were 6:3, untested; 8:4, 121; and 11:4, 117, respectively). Their task was to select from this array a group of pictures that 'are alike in some way'. The drawings represented familiar objects such as a pair of scissors, a doll, a garage, a bee, a pumpkin, various articles of clothing, a sail boat, a taxi, and so forth, and these were placed on a table in front of the child (Figure 6). He was first asked to identify each picture to ensure that he had seen and was familiar with all of them. Whenever he was unable to provide any sort of identification for a picture, he was told what it was. When he was asked to choose pictures that were alike in some way—'any way at all in which a group of things is the same'—and to remove them from the array. He could take as many pictures as he wanted. When he had completed his grouping, he was asked to tell how the pictures he selected were alike. The pictures were then replaced in their original

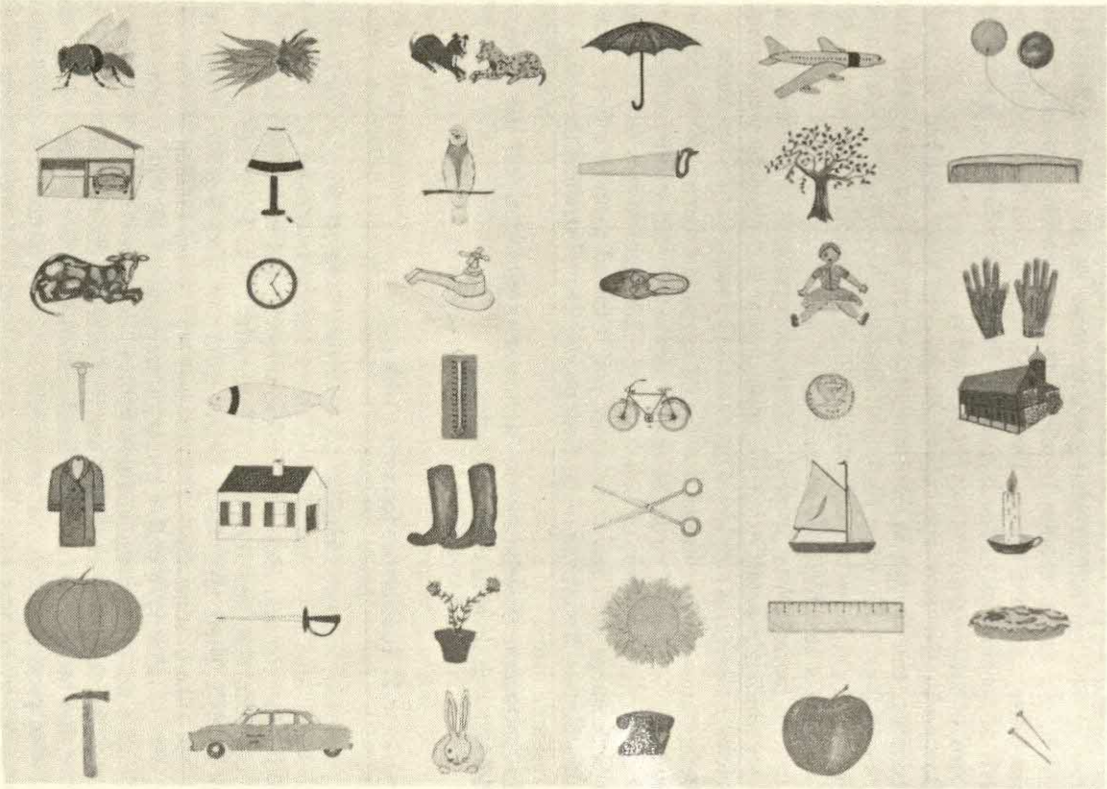


Fig. 6 Pictures (originals in colour) used in equivalence task with pictorial material

positions in the array, and he was asked to form another group. The task was repeated ten times, the child selecting new groups from the full set of pictures each time.

The picture-grouping task differs from the verbal task in several important ways that can be expected to influence the equivalence groups formed. Davidon (1952), in working with college students, has found that the nature of the materials affects the type of concept formed in the following manner:

> 1 When grouping verbal symbols there appears to be a greater tendency to attain concepts based on common *use* than when grouping pictorial symbols (drawings and photographs). And with pictorial symbols, conversely, there is a greater tendency to attain concepts based upon common parts.

Since in this task we are dealing with pictures rather than verbal materials, we might expect a greater use of perceptible attributes and a decreased use of functional characteristics. Because the task permits the choice of any number of objects for grouping, it might also permit more pair-formations based on idiosyncrasies shared by members of the pair. Both these characteristics should promote differences between the two procedures.

Consider first the relative use of perceptible and functional attributes as the basis of equivalence in the two tasks. As would be expected from Davidon's study, the picture task elicits a greater use of perceptible and a lesser use of functional attributes, as the basis of grouping than did the verbal materials. At age six, there is a significantly greater use of perceptible attributes as the basis of grouping in the picture task. Forty-seven per cent of the groupings in the picture task at this age are perceptually based, and 29 per cent in the verbal task. By age eleven there is still a difference, but it is not statistically significant (Figure 7).

Although picture materials produce a greater reliance at all ages on perceptible properties as the basis for judging likeness, the six-year-old still bases far more of his groupings on the way things look than do older children. The use of perceptible attributes declines steadily from 47 per cent at age six to 27 per cent at age eight to 20 per cent at age eleven.

What about the use of functional attributes as the basis of grouping? Functional equivalence is used to a significantly greater extent in the verbal task at age six and at age eleven. At age six, 49 per cent of the equivalence groupings are functionally based in the verbal task, while only 30 per cent are so based in the picture task. At age eleven, the difference is even greater, 75 per cent of groupings in the verbal task being based on functional attributes, but 47 per cent of the groupings in the picture task are so based.

And so, although there is an increased use of functional attributes as the basis of grouping, from 30 per cent at age six to 48 per cent at age eight, the use of function is less evident in the picture task than in the verbal task at

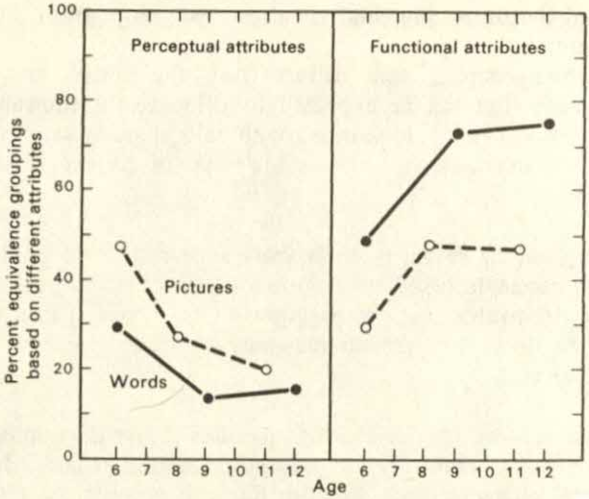


Fig. 7 Percent of equivalence groupings based on functional and perceptual attributes in verbal and picture tasks

all ages. It seems likely that in the picture task the child has a powerful alternative to functional grouping, for he can use nominal grouping readily in this array. Indeed, the use of nominally based equivalence increases steadily from 6 per cent at age six to 23 per cent at age eight, and to 32 per cent at age eleven. All the eleven-year-olds use nominal groupings at least once as the basis for equivalence, whereas 83 per cent of the eight-year-olds but only 47 per cent of the six-year-olds use class names as the basis of equivalence. In the verbal task, in which the items selected were not easily grouped by a conventional name, the use of nominal groupings remained fairly constant at about 10 per cent from age six to twelve.

In the picture task, then, growth expresses itself in a greater tendency to use the nominal technique of imposing equivalence. Although the young child knows the conventional groupings–even three-year-olds, when presented with the picture array and asked to pick out all the animals, all the articles of clothing, and so forth, can easily accomplish the task–their self-generated, self-directed use of linguistic terms for categorizing objects demonstrates a gradual course of development.

One wonders about the significance of nominal groupings in the picture series, whether it might not indeed be a verbal method of 'pointing' to pictures, a way of signifying them by label in a rather superficial way. 'Those are clothes', or 'Those are food'. It may well be that such superficial word labelling is tempted by pictures more than by other words. To say of words heard that they 'stand for food' seems somehow more banal than saying that some pictures 'are food'. Perhaps it is regarded by us all as a step ahead to describe a picture with a simple descriptive word, whereas no such sense of progress is felt in substituting one simple descriptive word for others. It may amount to some sort of culturally acknowledged superiority in having something pinned by a word.

As we might bet, the average number of items included in a group by six-year-olds are close to a pair. About 61 per cent of all groups formed by six-year-olds consists of pairs; this figure drops to 36 per cent for eight-year-olds, and still further to only 25 per cent of the groupings of eleven-year-olds. The pairings of six-year-olds are often capable of including 'logically' more than two items in the array, but they do not extend to these other items. The young child will form a group consisting of *house* and *barn* 'because they have red on them', and yet ignore the red *apple* and red *balloon*. Or he will select just two animals or two foods and fail to include the other animals or foods. Indeed, he may later form another pair on the same basis as an earlier two-item group. This finding suggests a comparable one by Vurpillot and Zoberman (1965) concerning the young child's method of searching for similarities. Vurpillot and Zoberman found that, in general, the young child is satisfied with a judgment of similarity when a single feature in two displays is found to correspond. The young children simply form 'perceptible' pairs, group on the basis of a single attribute, and have done with the task once a pair has been found.

Many of the pair groups formed by the younger children are sentential in nature–the grouping of two items in a sentence–'the *bunny* ate the *carrots*' or 'the *bee* stung the *cow*'. Such sentential structures, like the edge matching or chain complex, exhibit constraints running sequentially from one item to the next, but they lack the higher level summarizing constraint of the superordinate structure. If additional items are added to a group that begins as a sentential structure, there are few limits to what may be included. For example, '*Bee* can be around *cow, cow* is around a *person, person* has with him *apple, sword, rabbit,* and little *dog* he found, he got a ride back in the *taxi* to the *garage,* the *garage* has the other things (*clock, lamp, plant, thermometer*)'.

The use of sentential structures declines from age six on: six-year-olds construct 31 per cent of their groupings as sentential structures, eight-year-olds 20 per cent, and eleven-year-olds only 8 per cent. It would seem that for the young child the sequential form of the sentence is a powerful determinant of connections between things, while for older children other aspects of language become more instrumental in thought–for example, form class. This shift in equivalence is consistent with the transition from the syntagmatic to the paradigmatic responses in word association noted by Ervin (1961) and by Brown and Berko (1960) from grades 1 to 6.

In addition to the formation of sentential pairs at age six, the six-year-olds form complexive structures more often than do older children. Complexive structures decrease from 38 per cent at age six, to 18 per cent at age eight, to 12 per cent at age eleven. All the types of complexive structures noted in the verbal task also occur when the child is free to select the items to be grouped from an array of pictures:

Collections: The grouping of *boat, ruler, quarter, doll, bicycle, scissors, saw, shoe, gloves, barn, candle, pie, nails,* and *taxi* as, 'Some are red, some are gold, and some are yellow. One is white, some are brown, and some are blue'.

Edge matchings: Candle and *clock* are on a table and this (*lamp*) is round and this (*clock*) is'.

Associations: The grouping of *boots, cow,* and *gloves* as, 'These could be leather *gloves* and leather *boots,* and you get leather from *cows*'.

Key rings: 'You build a house with the *hammer, nails, screw,* the *barn* is next to the *house.* By the *house* you have a *tree*. You eat the *carrots, pie, apple, pumpkin* in the *house.* And you have *clock, bird, faucet* in the *house.*'

Multiple groupings: *Screw, ruler, nails, candle, hammer, taxi, coat, scissors, sword,* and *bicycle* are alike because, 'They have a part that you get dressed with, or they have holes in them, or you use them for tools, taxi goes with bicycle'.

As a corollary to the decrease in sentential and complexive groupings, it comes as no surprise to us that there is also an increase in superordinate constructions from 34 per cent at age six to 69 per cent at age eight to 85 per cent at age eleven. Increasingly with development the child isolates one or more attributes that are common to all the items in the group: 'They are all tools', or 'You can eat them', or 'They can all move', and so on.

With pictures as with words, the child first achieves invariance across a series of items by taking himself as a reference point. Eight-year-olds reflect this self-reference by using 'you' and 'I' in their responses more than either the younger or older children do. 'You use them for tools', or 'You wear them', or 'I can ride in them', are frequent in their protocols (see Figure 2), but the tendency to use such egocentrism as a crutch is less with pictures than with words.

To summarize: the same pattern of growth emerges whether we use words or pictures as stimuli, and whether the child is given items in a fixed order or chooses his own groups *ad libitum.* Equivalence for the six-year-old reflects a basis in imagery, both in what he uses as a basis for grouping and in how he forms his groups. From age six on, linguistic structures increasingly guide what and how things will be judged alike. With the development of symbolic representation, the child is freed from dependence upon moment-to-moment variation in perceptual vividness and is able to keep the basis of equivalence invariant. A first step away from domination by the perceptually salient comes when the child, at about age nine, takes himself egocentrically at a reference point for establishing equivalence among things. He does this by imposing upon the world what *he* can do to things, producing equivalence by reference to his own actions. In time, he accommodates to more conventional definitions of how things are 'alike'.

However, it is not entirely a matter of what features of the world serve as a basis for defining how things are alike. There is also the grammar of grouping what is alike and segregating it from the rest of the world. The effort begins with loosely ordered definitions of likeness, ordered sequentially and in a fashion that makes discrimination difficult and combination awkward. Only gradually and by dint of much language development does the child work his way toward true conceptual grouping, based on the rule of the superordinate class, which opens new possibilities for relating and combining and structuring information about the world. By early adolescence the rules are mastered, and then the question remains as to how well the child learns to apply them in a variety of contexts.

Let us now give a brief interpretation of the phenomena reported in the experiments in this article. [...] It is quite clear that the shift from complexive-perceptual definitions of equivalence to ones that are superordinate-functional is *not* a universal property of 'growing up'. For one thing, we see that the 'natural' terminus of growth depends to a very considerable extent on the pattern imposed by the culture. The techniques used in this article have, in modified form, been used in studies of children in Alaska, Mexico, and Senegal, and it is plain that school children in Dakar or Mexico City look very much like the school children of this article. But it is equally plain that the village child of rural Mexico and the unschooled Wolof of Senegal seem very different, much more complexive, much more perceptually oriented.

A second consideration has to do with the relation of the two systems of representation and their interaction in growth. We know from the searching studies of Wallach and Kogan (1965) that the two orientations–call them ikonic and symbolic as a shorthand–can be quite independent one of the other. That is to say, one finds children who are clearly proficient both in kinds of complexive and metaphoric activities of the first mode and in the more abstract activities of the second. In others, the first suffers a kind of replacement by the second, and a different style emerges. All this is within the limits of our own culture.

In conclusion, we would say that our middle-class children do show a pattern of growth in which 'appropriate' rational classification based on the functional properties of things 'replaces' earlier complexive grouping based on the surface properties of events. But such growth is not inevitable, not complete, and not something that invades every corner of the mind.

References

BROWN, R. W. and BERKO, J. (1960) 'Word association and the acquisition of grammer', *Child Development, 31,* 1–14.

DAVIDON, R. S. (1952) 'The effects of symbols, shift and manipulation upon the number of concepts obtained', *J. exp. Psychol., 44,* 70–80.

ERVIN, S. M. (1961) 'Changes with age in the verbal determinants of word association', *Am. J. Psychol., 74,* 361–72.

VURPILLOT, E. and ZOBERMAN, N. (1965) 'Rôles des indices communs et des indices distincts dans la différentiation perceptive', *Acta psychol., 24,* 49–67.

WALLACH, M. A. and KOGAN, N. (1965) *Modes of Thinking in Young Children.* New York: Holt, Rinehart and Winston.

WITTGENSTEIN, L. (1953) *Philosophical Investigations.* New York: Macmillan.

WITTGENSTEIN, L. (1958) *The Blue and Brown Books.* Oxford: Blackwell.

18 Provoked correspondence and equivalence of corresponding sets

J. Piaget and A. Szeminska

In analysing the beginnings of quantification, we find ourselves confronted with the problem of correspondence. To compare two quantities is indeed either to compare their dimensions, or to make a one-one correspondence between the elements. As a result of the work of Cantor, the second of these processes has been seen to be fundamental to the construction of the integer, since it provides the simplest and most direct measurement of the equivalence of two sets. The fact that the discovery of this operation came so late is due, as Brunschvieg clearly proved, to its very primitive character. Both counting on the fingers and the exchange of one object for another are indications of the considerable part played by correspondence in the synthesis of number.

Yet although one-one correspondence is obviously the tool used by the mind in comparing two sets, it is not adequate, in its original form or forms, to give true equivalence to the corresponding sets, i.e. to give each set the same cardinal value, which is seen as constant as a result of the correspondence. Either the correspondence is held in check by perceptual factors, or the correspondence itself develops from mere global correspondence between configurations into truly quantifying correspondence, which issues in necessary equivalence and thus in cardinal invariance. We shall now proceed to examine the question of the possibility of such development of correspondence.

We must first make a distinction between two kinds of situations in which the child is led to discover or to make the one-one correspondence. On the one hand, there are those cases in which he is required to assess the value of a given set of objects by comparing them with objects of the same kind. For instance, if two children are playing marbles, and one puts four or six on the ground, his partner will want to put one opposite each of them and thus obtain an equivalent set, without needing to be able to count. This correspondence between like objects raises the whole problem of cardination [which will not be discussed in this article].

On the other hand, there is an even simpler situation, which we shall first study, namely, correspondence between objects that are heterogeneous, but qualitatively complementary, a correspondence which is thus, as it were, provoked by external circumstances. During a meal, for instance, the child might be asked to put an egg in each egg-cup, to put a glass opposite each

---

Source: PIAGET, J. and SZEMINSKA, A. (1952) *The Child's Concept of Number.* London: Routledge and Kegan Paul, 41–56 (abridged).

bottle, or a flower in each of the little vases, etc. We must also include in this category the exchange for one object for another, for example one flower or one sweet for each penny, etc. In this article, our investigation will be restricted to situations in which the correspondence is provoked, with a view to establishing whether the one-one correspondence made by the child, or with his help, necessarily entails in his mind the idea of lasting equivalence between the corresponding sets. We shall in fact find that this is not so, but we cannot be too careful in our choice of examples, and shall therefore confine ourselves to the analysis of these elementary behaviours. [. . .]

The procedure will be as follows. We shall first examine the correspondence between six to ten glasses and bottles. Since this experiment does not give rise to the notion of lasting equivalence, we shall go on in §2 to the analysis of the correspondence between flowers and vases, which would seem to be even simpler, since the flowers are actually put into the vases and not merely near to them. The result being the same as before, we shall check the position by an even simpler correlation: that of eggs and egg-cups. It is simpler because each egg-cup can contain only one egg, whereas the relation between the number of glasses and bottles, or flowers and vases, is arbitrary. [. . .]

§1. One-one correspondence between glasses and bottles

Six little bottles (about one inch high, of the kind used in dolls' games) are put on the table, and the child is shown a set of glasses on a tray: 'Look at these little bottles. What shall we need if we want to drink?–*Glasses.*–Well, there they are. Take off the tray just enough glasses, the same number as there are bottles, one for each bottle'. The child himself makes the correspondence, putting one glass in front of each bottle. If he takes too many or too few, he is asked: 'Do you think they're the same?' until it is clear that he can do no more. Mistakes occur in fact only with children of the first stage (4–5 years). The correspondence can be made easier by getting the child to empty the bottles into the glasses, each bottle just filling one glass. Once the correspondence is established, the six glasses are grouped together and the child is again asked: 'Are there as many glasses as bottles?' If he says 'no', he is then asked: 'Where are there more?' and 'Why are there more there?' The glasses are then rearranged in a row and the bottles grouped together, the questions being repeated each time.

Stage I: No exact correspondence and no equivalence

> Bon (4:0): 'look at all these little bottles. What shall we need if we want to drink?–*Some glasses.*–Well, there are a lot here (putting them on the

table). Now put out enough glasses for the bottles, just one for each.–(He took the twelve glasses, but put them close together, so that the six bottles made a rather longer row.)–Where are there more?–*There* (the bottles).–Well then, put one glass for each bottle.–(He made the twelve glasses into a row the same length as that of the six bottles.)–Are they the same?–*Yes.*–(The bottles were then put further apart.) Is there the same number of glasses and bottles?–*Yes* (but he spread out the glasses a little more.)–(The bottles were then put still further apart.)–*There are only a few here* (the twelve glasses), *and there* (the six bottles) *there are a lot.*'

Gol (4:0) began by pouring the contents of each bottle into a glass. When he came to the fourth bottle he suddenly saw that he could not make the six bottles correspond to the twelve glasses, and cried: '*There aren't many bottles.*–Then you can take some glasses away.–(He left seven glasses for six bottles, putting the glasses rather closer together.)–Is there the same number of glasses and bottles?–*Yes.*–(One glass was then put in front of each bottle, so that one could be seen to have no corresponding bottle.)–*We'll have to have another bottle.*–(He was given one.) Is it right now?–(He so arranged them that the first bottle corresponded to the second glass, and so on up to the seventh bottle, for which there was thus no glass.) *No, here there's a glass missing, and there there's a glass that hasn't a bottle.*–What do we need then?–*One bottle and one glass.*–(He was given them, but he put them opposite one another and never made the correct correspondence.)'

Car (5:2): 'Arrange them so that each bottle has its glass.–(He had taken all the glasses, so he removed some and left five. He tried to make these correspond to the six bottles by spacing them out so as to make a row the same length.) Is there the same number of glasses and bottles? *Yes.*–Exactly?–*Yes.*–(The six bottles were then moved closer together so that the two rows were no longer the same length.) Are they the same?–*No.*–Why?–*There aren't many bottles.*–Are there more glasses or more bottles?–*More glasses* (pushing them a little closer together.)–Is there the same number of glasses and bottles now?–*Yes.*–Why did you do that?–*Because that makes them less.*'

These reactions are those of an earlier stage than that of correspondence in the true sense, a stage at which evaluation takes place through global comparison of the length of the sets in question. The case of Car shows this very clearly. He thinks that a row of five glasses contains more elements than a row of six bottles that are closer together, but he also thinks that by closing up the first row 'it makes them less' and that it then becomes equivalent to the row of six bottles! It follows that the notion of lasting equivalence between two sets is impossible, since equivalence depends on variable factors such as the length of the rows.

Stage II: One-one correspondence, but without lasting equivalence of corresponding sets

The following examples are those of children who are perfectly capable of making without hesitation the one-one correspondence between bottles and glasses. But although they are certain that there are as many glasses as bottles when they can see the corresponding elements opposite one another, they cease to believe in the equivalence when the rows are no longer of equal length.

> Hoc (4:3): 'Look, imagine that these are bottles in a café. You are the waiter, and you have to take some glasses out of the cupboard. Each bottle must have a glass'. He put one glass opposite each bottle and ignored the other glasses. 'Is there the same number?–*Yes.*–(The bottles were then grouped together.) Is there the same number of glasses and bottles?–*No.*–Where are there more?–*There are more glasses.'* The bottles were put back, one opposite each glass, and the glasses were then grouped together. 'Is there the same number of glasses and bottles?–*No.*–Where are there more?–*More bottles.*–Why are there more bottles?–*Just because.*'
>
> Mog (4:4) estimated that he needed nine glasses for the six bottles, then made the one-one correspondence and removed the three that were left over, and said spontaneously: *'No, it wasn't the right number.*–And are they the same now?–*Yes.*–(The glasses were put closer together and the bottles spread out a little.) Is there the same number of glasses and bottles?–*No.*–Where are there more?–*There are more bottles.*'
>
> Gin (4:11): 'Take just enough glasses off this tray for the bottles, one for each.–(He took all the glasses.) Do you think it's the same number?–*No.*–Take away the extra ones then.–(He made the one-one correspondence merely by looking at them, and left six glasses on the tray without counting.) Are they the same?–*Yes.*–Well, now put them so that we can see if it's right.–(He placed them correctly opposite the bottles.) *There.*–Is there the same number?–*Yes.*–(The glasses were then grouped together.) Is there the same number of glasses and bottles?–*No.*–Where are there more?–*There are more bottles.*–Why?–*Because there are more here* (pointing to the row of six bottles.)–(The glasses were spread out and the bottles grouped together). Are they the same?–*No.*–Where are there more?–*Here* (the glasses).'
>
> Gal (5:1) made six glasses correspond to six bottles. The glasses were then grouped together: 'Is there the same number of glasses and bottles?–*No, it's bigger there* (the bottles) *and smaller here* (the glasses).–(The bottles were then grouped together and the glasses spread out.)–*Now there are more glasses.*–Why?–*Because the bottles are close together and the glasses are all spread out.*–Count the glasses.–1,2 . . .6.–Count the bottles.–1,2 . . .6.–They're the same then?–

Yes.–What made you say they weren't the same?–*It was because the bottles are very small.'*

Mül (5:3) made the correct correspondence between bottles and glasses, but only after estimating at a glance that he would need two extra glasses. 'Were they the same?–*No, there were too many glasses.*–And are they the same now?–*Yes, they're the same.*–(The glasses were then grouped together and the bottles spread out.)–Are they the same?–*No, because that's bigger.*–Can you count?–*Yes.*–How many glasses are there?–*Six.*–And how many bottles?–*Six.*–So there's the same number of glasses and bottles?–*There are more where it's bigger.'*

Os (5:10) at once made the correspondence: 'Is there the same number of glasses and bottles?–*Yes, I've counted them.*–(The glasses were then grouped together.)–Is there the same number of glasses and bottles?–*No.*–Why?–*Because there are a lot here* (bottles) *and only a few there.*–(The bottles were grouped together and the glasses spread out.) Are they the same now?–*No.* Why?–*Because here* (glasses) *it's a lot and there it's only a few.'*

Fu (5:9) poured the contents of the six bottles into six glasses and put the glasses in front of the empty bottles. 'Is there the same number of bottles and glasses?–*Yes.*–(The bottles were grouped together in front of the glasses.) Are they the same?–*No.*–Where are there more?–*There are more glasses.*–(The reverse process then took place.) And now?–*There are more bottles.*–What must we do to have the same number?–*We must spread out the glasses like this, no, we'll need some more glasses.*'

Fra (6:3). Same reaction: when the glasses were grouped together, there were more bottles, and vice versa: '*There are more because they're not spread out.*' When finally he was told: 'Make it so that they're the same', he reconstructed the one-one correspondence.

Such are the reactions of the second stage. The first point to notice is that all these children are capable of making the one-one correspondence, but that as soon as the visual correspondence is destroyed, the quantitative equivalence and even the qualitative correspondence no longer exists for the child. It is as though, for the child, quantity depended less on number (a notion which, if our hypothesis is correct, is still only verbal, although the child can count correctly) or on the one-one correspondence between the objects, than on the global appearance of the set, and in particular on the space occupied by it. Even Mül, for instance, who could count, thought that 'there are more where it's bigger', although he had counted that there were six glasses in the group and six bottles in the row.

It might be argued that the mistakes are due to lack of understanding of the words used. May it not be that the child does recognize that the number of bottles and glasses remains the same when one set is grouped together, and that when he says 'there are more' he is merely expressing the idea that the shape of the set has changed and the space it occupies is greater? It is precisely because of this possibility, and because of the difficulty of ensuring that there shall be no verbal misunderstanding we shall give a wide variety of examples and situations in the course of this article. By continuing our analysis in the light of further facts we shall be able to choose between these two interpretations.

Even at this point, however, it seems desirable to make the following observations. In the first place, while it is difficult to find terms for the expression of quantitative equivalence that can be easily understood by children of between four and six, we have on the other hand no proof that a child of five like Mül is using the words 'six glasses' or even 'six' in general, in the same sense as we ourselves do. All that we can see is that Mül can apply the first six numerals to six objects, that is to say he can make words correspond to glasses as well as glasses to bottles. But does that prove that this verbal enumeration expresses a better quantification, from the child's point of view, than the space occupied, or that the fact of attributing numbers to objects answers the question 'how many?' in a truly numerical sense? We are obviously in no position to assert that this is so, since it is possible that at this level the correspondence between numerals and objects is still purely verbal, and that the child has not yet acquired the notions necessary for the construction of number itself, i.e. permanence and equivalence of sets irrespective of the distribution of the elements of which they are composed. The argument concerning the language used thus has another side, and all that can safely be said is that there is a discrepancy between the labelling with numbers and the visual intuition.

In the second place, when the child expresses a quantitative change, he does not always merely say 'there's more' or 'there's less', which would suggest that he was making a purely spatial evaluation without taking into account the discontinuous quantities, but frequently he is quite explicit, 'there are more glasses', or 'there are more bottles' (cf. Hoc, Mog, Gin, Du, Fu, etc.). Os says 'there are a lot here and only a few there'. Gal, who, unlike Mül, is finally convinced of the equivalence of the two sets on discovering that they contain the same number of elements, makes the position quite clear. He begins by saying 'it's bigger here', which he then translates into 'there are more glasses', and goes on to say that when the bottles are grouped together they 'are very small'. This last statement can only mean that the child had expected that the actual quantity would decrease, and on finding that the number of elements remained the same, reconciled this permanence of the number 6 with the contraction of the space occupied, by reducing the size of the elements.

In the third place–and this seems to us the decisive argument–when the correct answer is arrived at, it is clear what the child was thinking earlier. We shall see that, at the third stage, the child discovers, and stated explicitly, that the fact of grouping together or spacing out the elements in no way affects their number. This achievement is characteristic of the higher level, whereas before this level is reached, spatial alterations appear to the child to affect the quantification of the element.

Stage III: One-one correspondence and lasting equivalence of corresponding sets

Here are two examples of correct answers, which will serve for comparison with the answers given earlier:

> Pel (5:6) began by putting five glasses opposite six bottles, then added one glass: 'Are they the same?–*Yes.*–And now (grouping the glasses together)?–*Yes, it's the same number of glasses.*–Why?–*That hasn't changed anything.*–And if they're like that (grouping the bottles together and spacing out the glasses)?–*Yes, it's the same.*'
>
> Lau (6:2) made six glasses correspond to six bottles. The glasses were then grouped together: 'Are they still the same?–*Yes, it's the same number of glasses. You've only put them close together, but it's still the same number.*–And now, are there more bottles (grouped) or more glasses (spaced out)?–*They're still the same. You've only put the bottles close together.*'

It is clear that for these children, once the sets have become equivalent through the one-one correspondence, they remain so, irrespective of the arrangement of their elements. The difference between this and the earlier stage is brought out mostly obviously by Lau, for whom the number of glasses remained the same because 'you've only put them close together'. These answers indicate that the quantities remain equivalent although the space occupied changes, which suggests that earlier on, the problem for the child was to know whether the number varied with the configuration. The operation of one-one correspondence is thus constituted, and triumphs over mere intuitive or optical comparison.

We can now proceed to examine the significance of the three stages we have just described, or at least to state our hypotheses, which will be tested in the experiments of the following section.

As far as the first stage is concerned, the situation is clear. The child confides himself to a kind of global comparison, without one-one correspondence, merely making a spatial evaluation (length of rows, etc.). The third stage also is perfectly clear: one-one correspondence and lasting equivalence between sets. The explanation of the second stage, therefore, seems to lie in the fact that it is the transition between the two others, and the reactions of the children at this level must be taken at their face-value, without attempting to translate their thoughts in terms of higher concepts. For them, there is quantitative equivalence between two sets when there is one-one correspondence, but this correspondence is perceptual, or intuitive, thus involving a perceived contact between the corresponding elements. In the examples we gave this perceived contact was visual, but it might equally well be auditory or tactual. Precisely because of this limitation, once the contact is no longer perceived, the correspondence ceases to exist for the child, and the only criterion for evaluation remaining to him is the global, spatial criterion of the previous stage. Mül, for instance, although he can count up to six, says 'there are more where it's bigger'.

What does the child who knows that there are six glasses and six bottles mean, then, by the expression 'there are more'? And more generally speaking, what does the child mean by expressions such as 'there are more glasses' and 'here there are a lot and there there are only a few'? It would be absurd to suggest that they imagine that the actual number of objects varies, since the basis of our whole interpretation is our belief that these children do not yet possess the notion of number. On the other hand, and for the same reason, the child cannot simply mean that the space has increased while the number remains the same. The only possible interpretation, therefore, is to assume a kind of lack of differentiation between the number and the space occupied. The child's evaluation is, as we have already said, global, and not as yet analytic, the only evaluation at his disposal being perceptual correspondence. Fu illustrates this very clearly when he says that in order to restore the correspondence between six glasses in a group and six bottles in a row, either the glasses must be spaced out, or more glasses must be added, as though the two solutions were equivalent.

We are thus confronted with two problems. The first is that of the transition from global quantification by means of perceptual relationships of length or space occupied, to intuitive one-one correspondence; the second is that of the change from this intuitive correspondence to operational correspondence with lasting equivalence. Further examples will, however, be necessary before these problems can profitably be discussed.

§2. Correspondence between flowers and vases, and between eggs and egg-cups

It seems obvious that the closer the one-one correspondence between the elements, the more lasting the equivalence of the corresponding sets will be. If, therefore, a flower is put into a vase, or an egg into an egg-cup, the link between the corresponding elements will be closer for the child than when a glass is merely put opposite a bottle. He will therefore have less difficulty in understanding that the quantity of flowers or eggs remains equal to that of vases or egg-cups when the flowers or eggs have been taken out and piled together.

This is an important situation from two points of view. In the first place, it provides further evidence of the correctness of our interpretations, for if the same children answer the same questions better when the correspondence is intuitively closer, they show that it is not a matter of verbal misunderstanding, but that the quantifying value of the correspondence is greater or less according to the content of the particular problem. In the second place, the difference in the degree of ease with which the questions are answered will enable us to analyse the answers better than if there were a general lack of understanding on the part of the child.

The technique used was as follows. In the case of flowers and vases, the child's interest was first aroused by a little game: 'What shall we put into these vases?–*Some flowers.*–Then we must go and get some from the garden, one flower for each vase, as many flowers as vases?' A number of flowers were then put before the child, more than there were vases, and we observed how he made the correspondence. He either put one flower opposite each vase, or made a more or less compact row the same length as the row of vases. He was then asked to check his result by putting one flower into each vase, thus obtaining a one-one correspondence. The flowers were then taken out and bunched together (or the vases grouped) and as before the child was asked whether there was still the same number of both. In the case of eggs and egg-cups, the technique was the same, but in addition the eggs were clustered first close to the egg-cups, and then some distance away, in order to test whether optical contact does in fact affect the estimate of equivalence.

Stage I: Global comparison without one-one correspondence or lasting equivalence

Here we have some examples of children who merely made a row of flowers the same length as the row of vases:

> Fum (4:4) began by taking the flowers one at a time, looking at each vase in turn, but he was only able to keep this up for a few units and then made a global estimate. 'Are they the same?–*Yes.*–Try them and see.–(He put the flowers into the vases and found that he was three short.) *I need some more flowers there* (adding them).–And now are they the same?–*Yes.*–Now we'll take the flowers out for a minute and change the water (closing up the vases and spacing out the flowers). Is there the same number of vases and flowers?–*There are more flowers.*–Try them.–(He spaced out the vases.) *No, they're the same.*–(The vases were again closed up.)–*There are more flowers.*–Why?–*Because there's a flower here* (pointing to a flower which wasn't opposite a vase).–Do you think all the flowers will go in?–*I think we'll have to take those away* (the two flowers that were beyond the end of the row of vases). *I'll put the flowers in quickly.* (He did so, and finding that he needed the two he had taken away he added them.)–Now we'll change the water again, shall we? (The flowers were again taken out and bunched together.) If we put the flowers back in the vases, will there be the same number or not?–*I think they're the same. No, there are too many vases.*–Well then, you do

> it yourself so that they're the same.–(He closed up the vases!) Do you think that's right?–*Yes, I think that's all right.*'
>
> Gui (4:4) put thirteen flowers close together in a row opposite ten vases rather more spaced out, although he had counted the vases from one to ten. Since the rows were the same length, he thought that the flowers and vases were '*the same*'. Then you can put the flowers into the vases?–*Yes.*' He did so, and found he had three flowers over. The flowers were taken out and bunched together in front of the vases. 'Is there the same number of vases and flowers?–*No.*–Where are there more?–*There are more vases.*–If we put the flowers back into the vases, will there be one flower in each vase?–*Yes.*–Why?–*Because there are enough.*–(The vases were closed up and the flowers spaced out.)–And now?–*There are more flowers.*'

Here are three examples of reactions at this stage to eggs and egg-cups, the most primitive of them showing (like Fum in the case of the flowers) that the child could not yet grasp even the idea that the original situation could be restored:

> Fra (4:3): 'Take just enough eggs for the egg-cups, not more and not less, one egg for each cup.–(The child made a row the same length but containing far too many eggs.) Is there the same number of eggs and egg-cups?–Yes.–Well then, put the eggs in, to see whether they're right.–(He did so.)–Were they the same?–*No.*–Are they the same now?–*Yes* (removing the extra eggs).–Now we'll take all the eggs out (making them into a heap in front of the egg-cups). Are they the same now?–*No.*–Why?–*There are more egg-cups.*–Are there enough eggs for the egg-cups?–*I don't know.*–(The egg-cups were put closer together and the eggs spread out.) Look, now is there the same number of eggs and egg-cups?–*No, there are more eggs.*–Are there enough egg cups for the eggs?–*No, I don't know.*
>
> Zu (4:9) also began by putting opposite the egg-cups a row of the same length but containing too many eggs. He then put the eggs in and removed the extra ones. He himself then took out the eggs and made a pile of them in front of the egg-cups. 'Is there the same number of eggs and egg-cups?–*No, there are a lot of egg-cups and less eggs.*–Are there enough eggs for the egg-cups?–*No.*' All the eggs were then removed and (for seven egg-cups) only four were put back, well spaced out. 'Are there enough eggs for these egg-cups?–*Yes* (the length of the rows was the same).–Put them in yourself, to see.–(He put them in and seemed very much surprised that there were not enough.)–And now, are they the same (having removed the four eggs and put opposite the seven egg-cups a row of equal length, but composed of twelve eggs).–*Yes.*–Exactly?–*Yes.*–If we put them into the egg-cups, will there be any over?–*No, they'll all go in.*–Try it.–(He was again very much surprised.) *There are*

some left over!' With only three eggs, widely spread, for seven egg-cups, Zu replied correctly that '*There'll be some empty egg-cups*,' but with 5 eggs spaced out he again thought that they would exactly correspond!

It is clear that these children were not capable, by themselves, of making the one-one correspondence, and that they would not have discovered it if they had not been forced to do so by the relations between container and content as exemplified by vases and flowers or egg-cups and eggs. As for the equivalence of the two sets, it is obviously entirely based on a perceptual comparison of the length of the rows, and as soon as the elements in one set are put closer together or further apart then those of the other, the equivalence is no longer recognized. Thus Fum, in spite of having put one flower into each vase, thinks that the one-one correspondence breaks down when the flowers are taken out of the vases and put further apart, and he even goes so far as to remove some flowers in order to restore the correspondence! Similarly Zu's estimates are so completely dependent on the space occupied that he attempts to put first four, then twelve and then five eggs into seven egg-cups, and yet estimates that the seven eggs that he himself had put in, then taken out and placed in a more compact row, could not be made to correspond to the seven egg-cups! These amazing reactions indicate how little the child differentiates during the first stage between the discontinuous quantity and the space occupied. At this level, even when the one-one correspondence has been established through force of circumstances, the child doubts, if the perceptual aspect of one set is changed, whether a return to the original position will restore the correspondence. When, like Gui, the child thinks it will do so, the reason may well be that he merely remembers the correspondence perceived earlier, and there is no proof that in the meantime he considered the equivalence to be lasting. He says, in fact, that 'there are more flowers' when the vases are put closer together, and conversely.

Stage II: Intuitive one-one correspondence without lasting equivalence

Children at this stage are more advanced than those at the first stage in that they can spontaneously make the one-one correspondence, but they are as incapable as the others of deducing that the correspondence involves lasting equivalence of the sets irrespective of the arrangement of the elements. Here we have some examples:

Dal (4;6), after examining the ten vases, took nine flowers, thinking he had seen the exact number he needed. When he came to the seventh vase he saw that he wouldn't have enough, and took another. When all the flowers had been put into the vases, they were taken out and bunched together: 'Is there the same number of flowers and vases?–*No.*–Why?

–There are more vases.–And now (reversing the process)?*–There are more flowers.*'

Sim (5:7) took six eggs to correspond to six egg-cups and put the eggs in. They were taken out and spaced further apart: 'Is there the same number of eggs and egg-cups?*–No.*–Where are there more?*–Here* (eggs).–If we wanted to put one egg back in each egg-cup would there still be the right number?*–Yes . . . I don't know.*'

Dum (5:8) also took 6 eggs to correspond to six egg-cups and put them in himself. When they were taken out and put in a pile in front of the egg-cups, Dum thought they weren't the same. 'Why?*–Because you did that* (making a gesture to indicate closing up).–Are there enough eggs for the egg-cups?*–No.*–(The egg-cups were put closer together and the eggs further apart.)–Are they all right now?*–No, because there are more eggs.*'

These few examples are sufficient to confirm that there is a second stage, which occurs between the stage of spontaneous non-correspondence and the stage of lasting equivalence. The child can by himself produce a one-one correspondence, but it is still intuitive, since there is no longer equivalence once the configuration of the set is changed. Moreover, although some children at this stage think that a return to the original position is possible they do not see it as necessary. Sim, for instance, who assumes it in the case of the flowers and vases, is less certain for the eggs and egg-cups. In addition to this, even when the child accepts the possibility of a return to the original position he does not draw the conclusion that the equivalence persists during the interval. It is again Sim who says that 'there are more flowers' than vases when the vases are close together, although he thinks that one flower could be put back in each vase, and says explicitly, when asked 'Are they both the same, then?', 'No, here there are more, because they're spread out'. There could be no better indication that, for the child of this level, quantification is reducible neither to number (most of these children can count up to ten), nor to one-one correspondence, but to an intuitive correspondence depending on the perceptual configuration of the sets.

Stage III: Operational correspondence and lasting equivalence

The interest of these experiments lies in the fact that, being slightly simpler than those involving glasses and bottles, they provide us with examples of an intermediary nature, which facilitate the analysis of the method by which the child finds the correct solution. Here are some examples:

Du (5:8) was still at the second stage as far as the experiment with glasses and bottles was concerned. He made the correct correspondence between flowers and vases, and when the flowers were taken out and

> bunched together, he began by saying '*there are more vases*'. When the opposite was done, he said '*there are more flowers*'. He was then given another bunch of flowers of a different colour: 'Put one in each vase.–(When he had done so, they were taken out and bunched together in front of the vases.) And now, is there still the same number of flowers and vases?–*Yes.*–Why?–*Because we'd put them all there* (in the vases).' When, however, the vases were put closer together and the flowers spaced out some distance away, he again fell into his original mistake.
>
> Mou (5:8) also said that there were still as many vases as flowers when the flowers were bunched together. But when the vases were together he thought they were more numerous: '*There are more.*'
>
> Os (5:10) was also at the second stage as far as glasses and bottles were concerned. In the case of the flowers, he wavered between the solutions of the second stage and those of the third. He began by counting ten vases, and then ten flowers, putting them one by one into the vases as he counted. They were then taken out and bunched together near the vases: 'Is there the same number?–*Yes, because there are ten* (vases) *and here* (flowers) *there are ten.*–(The vases were closed up and the flowers spaced out some distance away.) And now?–*No, here* (vases) *there are only a few.*–Well look, here are some pink flowers. Take the right number for the vases (which were again in a row).–(He counted quietly as he put them into the vases.)–(They were then removed and spread out on the other side of the vases, the blue flowers remaining near the child.)–Is there the same number of pink flowers and blue flowers?–*Yes, there are ten here and ten there.*–And is there the same number of pink flowers and vases?–*No.*'

Here are a few reactions of the same type with respect to eggs and egg-cups. (It is to be noted that the child sometimes assumes equivalence when one set is grouped in close proximity to the other, but that this assumption diminishes with distance!)

> Gal (5:1) whose reactions at the second stage we have already seen in §1, at once took seven eggs to correspond to seven egg-cups. When the eggs were removed and grouped in front of the egg-cups, he still thought they were equivalent. 'Why do you say they're the same?–*Because.*–(The eggs were spread out some distance away.) Are they still the same?–*No.*–Why?–*Because there they're spread out and here they're close together.*–But if we put them back, would there be the same number of eggs and egg-cups?–*Yes.*'
>
> Os (5:10) counted the same number of eggs and egg-cups. The eggs were taken out and grouped together in front of the egg-cups: 'Is there the same number?–*Yes.*–(The eggs were put some distance away.) Are they still the same?–*No.*–Where are there more?–*There are more eggs.*–Would all the eggs go into the egg-cups?–*Yes.*'

The significance of these intermediary reactions is obvious. Generally speaking, they mark the beginnings of the construction of operational correspondence as distinct from optical or intuitive correspondence. Du, for instance, after saying that the sets were not equivalent when the one-one correspondence was no longer visible, was able to grasp that the number of flowers bunched near the vases was the same as the number of vases, for the very good reason that 'we'd put them all there', i.e. that the flowers had earlier been contained in the vases. But he was not capable of applying the same reasoning when the same flowers were spaced out, and thought they were more numerous because they were further away. Mou thought that the quantity of flowers remained constant, but that there were more vases when they were grouped together, thus using the criterion of density rather than of space occupied. The reaction of Os is extraordinary, in that he assumed that there was an equal number of vases and flowers when the flowers were bunched together but not when they were spaced out, and that there was an equal number of pink and blue flowers when both were spaced out but not as many pink flowers as vases when the vases were grouped and the flowers spaced out. Obviously the explanation is that when the flowers were bunched together near the vases they reminded him that they had all been in the vases, whereas when they were spaced out, and there was no optical contact, he forgot this. This is confirmed in the experiment with eggs and egg-cups. Os thinks that they are equivalent when the eggs are clustered near the egg-cups, but not when they are some distance away, and Gal shows clearly that when he is thinking of the possibility of putting back the eggs into the egg-cups he still believes in the equivalence, but ceases to do so when the eggs are further away. In a word, the child is beginning to free himself from perception and to construct a correspondence with equivalence that is truly intellectual. When the eggs are grouped together, but remain close to the egg-cups, the child assumes equivalence because the optical contact is a sufficient reminder of the correspondence, although the configuration has been changed. But when the eggs are widely spaced, thus being some distance from the egg-cups, the equivalence is no longer recognized, because the operation of correspondence is not yet completely freed from perception.

In the case of children who definitely reached the third stage, however, the operation is no longer affected by intuition and these children therefore grasp both reversibility and equivalence.

Fet (5:5) arranged ten flowers in a row in front of the vases and then put them in. They were taken out and bunched together: 'Is there still the same number of both?—*Yes.*—And if we put them like that (widely spaced some distance away)?—*Yes.*—Why?—*Because they were inside the vases.*'

Bet (5:8). After he had made the correspondence without counting, the flowers were bunched together some distance away from the vases.

'Is there the same number?–*Yes.*–Why?–*Because they go in* (i.e. into the vases).

And with the eggs grouped together in front of the egg-cups: 'Is there the same number?–*Yes.*–Why?–*Because they're like that* (making the gesture of grouping together).–And now (with eggs spaced out and egg-cups close together)?–*Yes.*–Why?–*If you spread the eggs out, it's the same number.*'

Pit (6:11). Same reactions. When the eggs were spread out, they were still the same '*because they all go into the egg-cups.*'

On reading these very simple answers, one wonders how it can have taken the child so long to grasp the lasting equivalence of corresponding sets. Yet the difference between these children and those quoted earlier is an essential one, indicating the triumph of the operation properly so called over perception.

Hitherto, the only quantification of which the child was capable was based on modification of a perceptual and spatial character, and the one-one correspondence itself was not quantifying. In other words, in the first stages, the qualities perceived by the child give rise only to simple quantitative relationships (more or less 'big', 'long', 'small', 'narrow', etc.) without operations in the true sense. These qualities are, indeed, not coordinated or multiplied one with another: the child does not see, for instance, that if the elements of a row are spaced out the number per unit of length diminishes, and that if they are closed up, this relative number increases. Thus during the first stage, the child judges the quantity only by the length of the row, without constructing even intuitive correspondences. During the second stage, the child has become capable of coordination of relationships, but only on the intuitive plane, i.e. he can make the correspondence when the elements of the two sets are opposite one another, but he no longer assumes equivalence when there is a change in the distribution of the elements of one set. The reason for this is that quantifying correspondence presupposes, in addition to merely perceptual correspondence, even if it is qualitatively exact, a more advanced operation, namely the equalization of differences, i.e. a coordination of the displacements such as they can offset one another when they become reversible. As long as the child is incapable of this multiplication, which is mathematical, and no longer merely qualitative, the correspondences he makes do not lead to lasting equivalence. That is why, even when the children who think that the bunched flowers are less numerous than the vases to which they correspond, assume that they could be put back one into each vase, it is not a case of a logically reversible operation, but only a foreseeing an empirical return to the original position. During the third stage, it is precisely this multiplication that takes place, because the child discovers that any spatial modification in the distribution of the elements can be corrected by an inverse operation. This is expressed by Fet and Bet when they give as reasons for their belief that the two sets remain the same, 'because they were inside the vases', 'because they go in', 'if you spread the eggs out, it's the same number', and 'because they all go into the egg-cups'. These reasons, which have no value for the children of the earlier stages, only become valid when reversibility is understood, and understood as a source of equivalence. It is therefore clear that the triumph of the operation over perceptual intuition is the outcome of the progressive reversibility of thought. Perception is essentially irreversible, but as it progressively resolves itself into judgements of relations, the reversible operations which result are able to dominate it, and thus to replace the intuitive correspondence by a quantifying operational correspondence which guarantees, in spite of appearances, the necessary and lasting equivalence of corresponding sets.

19 The understanding of invariance by very young children

P. E. Bryant

A simple and yet extremely important question in developmental psychology is whether young children understand that a quantity remains invariant when it is perceptually transformed. The question is whether a child who sees, for example, someone spreading out a row of counters, realizes that their number stays the same over the transformation. It was Piaget and his colleagues who originally posed this question (Piaget and Szeminska, 1941) and their answer was that children do not understand the principle of invariance until they have reached the age of, approximately, eight years. This claim was based on the results of a series of experiments, called conservation experiments, which showed that children younger than eight tended to change their judgments about a quantity when it was transformed.

A typical conservation experiment comprises two stages. For example, the child is asked in the first stage to compare two perceptually identical rows of counters placed alongside one another. Then in the second stage the experimenter transforms one row by spreading it out (or bunching it up), and the child is again asked to compare the two rows. Younger children usually say that the rows are equal in the first stage, and then after the transformation judge the longer of the two rows to be more numerous than the shorter. Older children, on the other hand, maintain their judgment that the rows are equal over the transformation. This is a result which has been replicated many times (Sigel and Hooper, 1968).

Piaget argues that the inconsistency of the younger children's judgments demonstrates that they cannot understand that the number of counters is unaffected by the perceptual transformation. There is, however, an alternative explanation, which is that the young children do understand invariance, but that their inconsistency simply represents the resolution of a conflict between two imcompatible hypotheses.

There is no doubt that children who cannot yet count or use numbers effectively can nevertheless form consistent hypotheses about quantities which they are asked to compare. Their behaviour is often far from random when they are simply asked to compare two rows of objects. With some displays their judgments are often uniform, and consistent over time, and this must mean that they have consistent hypotheses about the quantities concerned. If, for example, three-year-old children are given two identical

---

Source: *Canadian Journal of Psychology/Review of Canadian Psychology*, 1972, *26*, 78–96.

rows parallel to each other, and arranged like two ranks of soldiers, they are very well able to judge that the rows are equal in number. Here they seem to be using the one-to-one correspondence cue, which involves noticing that each of the objects in one row has its obvious counterpart in the other, and vice versa. This, of course, is a correct cue.

Another rather different example of a consistent quantity judgment at this age is young children's use of the length cue. If they are given two rows, unlike in perceptual appearance and differing in length, they consistently judge the longer row as the more numerous one (Binet, 1890; Piaget, 1968). The length cue is, of course, an incorrect cue, since the more numerous row may very well be the shorter one. However, they seem to use both cues as consistently as each other. The only difference between the two is that the former must always be right, while the latter must often be wrong.

Thus what is distinctive about the young child's quantity judgments is that they are sometimes incorrect. This must mean that, although these children are able to produce consistent quantity hypotheses, they do not yet understand what constitutes a sufficient basis for such an hypothesis.

Applying this analysis to the number conservation problem it can be seen that the experiment involves a transformation from a display which normally produces a judgment that the rows are equal, into a display which normally produces a contrary judgment that they are unequal. If we assume for the moment that the child does understand invariance, we can suggest that the transformation faces him with a conflict between two judgments, the first that the rows are equal, the second that they are not. He can only adopt one of these two incompatible judgments and realizes that one of them must be wrong. However, he has no systematic way of working out which is the better of the two. Faced with this dilemma he falls back on the more recent of the two. His inconsistent responding is the result not of a misunderstanding of invariance, but of the arbitrary resolution of a conflict between past and present judgments.

This explanation seems as cogent as Piaget's, and it is certainly not ruled out by any of his results. It follows that the traditional conservation experiment does not constitute a proper test of the understanding of invariance. A far better test would be a transformation which definitely avoids the possibility of such a conflict.

The solution is to find displays about which young children appear to have no consistent hypotheses at all, displays with which their behaviour is quite random. It is then possible to transform a display which does produce a consistent judgment into one of these displays which does not. This transformation avoids hypothesis conflict, and yet tests invariance. Children who do not understand invariance should respond systematically before the transformation, and then, after it, revert to random chance level behaviour. Without the invariance principle they could have no possible reason for maintaining a response over the transformation. Moreover, if they do transfer the systematic response over the transformation to a display which on its own would not have produced this response, they are demonstrably using the principle of invariance. The aim of the experiments reported below was to institute such a transformation, and to compare performance on it to performance on more traditional kinds of transformations.

Experiment I (A-B, A-C)

Method

Subjects. The *S*s were children at four age levels. The age groups were:

(*a*) 3-year-olds (mean age 3 yrs 9 mths: range 3 yrs 6 mths–3 yrs 11 mths);
(*b*) 4-year-olds (mean age 4 yrs 8 mths: range 4 yrs 3 mths–4 yrs 11 mths);
(*c*) 5-year-olds (mean age 5 yrs 6 mths: range 5 yrs 4 mths–5 yrs 8 mths);
(*d*) 6-year-olds (mean age 6 yrs 6 mths: range 6 yrs 3 mths–6 yrs 8 mths).

There were fifteen *S*s to each group. The first two groups were seen in two nursery schools and the second two in two infant schools in Oxford.

Displays. The displays were made up of coloured counters, placed on a square 2 ft × 2 ft broad. Each display consisted of two parallel rows of counters.

The first problem was to find displays which, presented on their own, produce inconsistent chance level responding. It is very difficult to calculate what is a chance level display, when *S*'s task is to judge whether two rows are equal or not, as it is in the conservation problem. The child could very well have unequal response biases for 'same' and 'different', which are quite independent of the displays themselves. This would mean that he could apparently produce consistent judgments with displays about which he had no hypothesis at all. Moreover, this kind of bias, which is very difficult to measure, might change through the experiment, and might also vary between children.

The solution is to present two rows with unequal numbers of counters, and to ask *S* which is the more numerous row. If *S*s have no hypothesis about which row is the more numerous their performance can only remain around the 50 per cent success level. This was the procedure adopted in the experiments.

This leaves the question of the actual number of counters to be used. Pilot studies showed that it is very difficult to find a display which produces chance level responding when small numbers are used. On the other hand, it is quite easy to find displays with large numbers of counters which do produce chance level responding in young children. The distinction seems to be between displays with rows of less than about eight counters and displays with rows of above that number. This may well be a 'subitising' effect (Woodworth and Schlosberg, 1954). Therefore, in order to ensure chance level responding in some displays, large number displays were used throughout. All the displays contained one row with twenty counters, and another with nineteen.

There were three kinds of display, and they are illustrated in Figure 1.

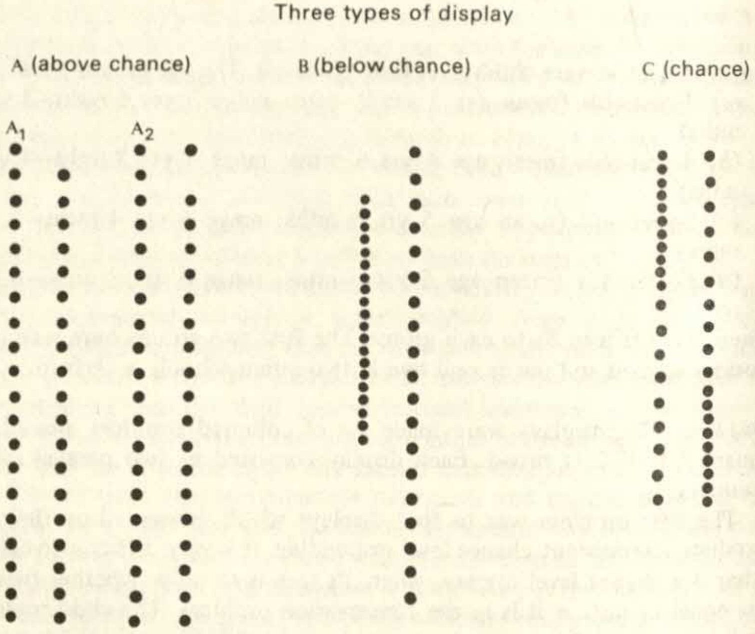


Fig. 1 Above chance, below chance, and chance displays

They differed from each other in their level of difficulty. Pilot experiments, whose results were in the experiment itself, established that different displays, presented on their own, produced performance varying from well above to well below chance level.

The displays were:

1 A (*above chance*). Given A_1 or A_2 on their own young children can consistently judge correctly which of the two rows is more numerous. In both displays each counter has its obvious matched counterpart in the other row, except that one counter in the more numerous row has no match. The intervals between counters is 1 inch in both rows in one version of these displays, and ½ inch in another.

2 B (*below chance*). In this display the longer row is the less numerous one.

Given the B display on its own children of six years and younger tend to judge consistently and incorrectly that the longer row is more numerous. Here the counters in each row do not have their obvious counterpart in the other row. The gaps between counters in the longer row are 1 inch and in the shorter (more numerous) row ½ inch.

3 C (*chance*). This display presented on its own produces chance levels performance in children aged between three and six years. They cannot judge which of the two is more numerous, and they have no consistent bias for either row. The counters in each row do not have their obvious counterparts, and the rows do not differ in length. When neither of these two cues is available, young children appear to be at a loss. In the row with twenty counters there were nine 1-inch intervals, and ten ½-inch intervals, while in the row with nineteen counters there were ten 1-inch intervals and eight ½-inch intervals between counters.

In every display the counters were always one colour in one row and another in the second. The colours used (red, green, blue, and yellow) were systematically varied. The rows stretched away from *S* and thus were an equal distance from him one 1½ inch to the right, the other 1½ inch to the left of the midpoint of the board.

The experimenter (*E*) sat to one side of the board, and this side was marked with 1-inch and ½-inch intervals, to help *E* arrange the counters. The marks could not be seen by *S*.

Procedure. Each *S* came to three sessions, a pre-test, an experimental session and a post-test, in that order. The purpose of the *pre-test session* was to establish baseline levels of performance with each display presented on its own. Once these baselines are established it is possible to measure whether information is transferred over a transformation, by comparing the baseline performance on a display with performance on the same display following transformation. If the two differ it must be because of information carried over the transformation.

For this reason each trial in the *experimental session* involved a transformation from one display to another, and the purpose of this session was to measure whether information could be carried across a transformation.

The procedure in the *post-test session* was identical to that in the pre-test. Its purpose was to check that any differences that did occur between pre-test and experimental session performance were not simply due to learning which had taken place between those two sessions.

Pre-test and post-test procedure. The procedure was exactly the same in these two sessions. There were twenty-four trial sessions. Each trial consisted of a display presented on its own. Each of the three displays (A, B, and C) was presented in eight trials, the order of these trials being randomised and varied between *S*s. The A_1 and A_2 displays were each presented in half the A trials.

In each trial *E* set out the two rows and simply asked *S* to point out the row 'with more counters'. As soon as the judgment was made *E* removed the counters completely, and set out the display for the next trial. Thus no transformation was involved. *S* was never told whether he was right or wrong.

Experimental session procedure. This session consisted of sixteen trials, in each of which one display was transformed in full view of *S* into another display. The transformation was always from the easy display (A) into one of the more difficult displays (B or C).

Thus there were two transformations, A-B and A-C. In the A-B trials *S* was first shown display A and was simply told to look at it for approximately 15 seconds. Then *E* transformed the A display into the B display. *S* was then asked which row had more counters. The procedure was exactly the same in the A-C trials except that the second display was the C one.

There were eight A-B and eight A-C trials. The order of the trials, and the position of the more and less numerous rows were systematically varied between trials in two Gellerman series (Gellerman, 1933). Also, on half the trials the two rows preserved the same position through the trial, while on the other half the position of the two rows was reversed during the transformation, the row on the left being moved to the right and vice versa. A_1 and A_2 displays were each used on half the trials.

Summary of procedure. The experiment thus involved three sessions and three displays. The pre- and post-tests established baseline levels for the three displays. In the experimental session the above-chance display (A) was transformed into either the below-chance (B) or the chance (C) displays. Thus the transformation in A-B trials was from a display which produces one hypothesis into a display which produces a contrary hypothesis. In A-C trials the transformation was from a display which produces a consistent hypothesis into a display which produces no consistent hypothesis at all. If in either the A-B trials or the A-C trials *S*s make more correct responses with the B or C displays than they do in the baseline pre- and post-tests, they will demonstrably have transferred information about quantity across a perceptual transformation.

Results

The mean scores for the three sessions are shown in Table 1.

Turning first to the baseline pre- and post-test scores, it can be seen that these amply confirm the pilot results. The performance of all age groups was well above chance on A, well below chance on B, and never differed significantly from chance on C.

The test for invariance is the experimental session scores. Young children make the correct judgment in A, the pre-transformation display. If they are able to maintain this hypothesis across transformation to B or C

Table 1 *First experiment: mean scores on* A, B, *and* C *displays (*A – B, A – C*)*

			Pre-test	*Exp. Session*	*Post-test*				*Pre-test*	*Exp. Session*	*Post-test*
Three-year group	A	*X*	7.13		7.40	Five-year group	A	*X*	7.27		7.40
		SD	0.88		0.88			*SD*	0.68		0.31
	B	*X*	1.40	0.90	0.80		B	*X*	0.66	1.53	0.53
		SD	2.06	1.26	1.38			*SD*	1.30	1.22	1.02
	C	*X*	4.1	6.7	3.5		C	*X*	4.40	6.53	3.80
		SD	1.26	1.53	1.45			*SD*	1.89	1.09	1.25
Four-year group	A	*X*	7.07		7.33	Six-year group	A	*X*	7.33		7.47
		SD	1.06		0.69			*SD*	0.69		0.72
	B	*X*	1.6	1.2	1.7		B	*X*	0.87	1.27	1.20
		SD	1.82	1.83	1.73			*SD*	1.75	2.14	1.68
	C	*X*	3.4	7.0	4.1		C	*X*	4.20	7.27	3.67
		SD	1.93	0.97	1.37			*SD*	1.45	1.06	1.40

(Exp. session scores are for B or C after transformation from A.)

performance on B or C should be better in the experimental session than in the two baseline sessions.

Table 1 shows that this kind of transfer did take place with C but not with B. Performance on C was at chance-level when the display was presented on its own, and yet rose to well above chance after the transformation. This pattern was consistent for all age groups. So even three-year-old children are able to transfer a quantity judgment across a transformation to a display which would not on its own have produced this judgment. To do this they must understand, and be using, the principle of invariance.

No such improvement occurred with the B display. Here performance remained at the same low level in all three sessions. Baseline scores demonstrate that B presented on its own produces the opposite judgment to A presented on its own. It can therefore be argued that when children do change judgments they do so as the result of a conflict between opposite hypotheses.

Performance on the B and C displays was subjected to a three-way analysis of variance in which the main terms were Ages, Displays, and Sessions. The Sessions term tests whether there is a difference between Pre- and Post-test and Experimental sessions. The main results of this analysis were a significant Display term (F 304.407, df 1,56, $p < 0.001$) which showed that the C display was generally easier than the B display, a significant Sessions term which was due to better performance in the Experimental Session, and a significant Displays × Sessions interaction (F 24,193, df 2,112, $p < 0.001$). Tukey tests showed that the cause of this interaction was that performance on C was better in the experimental than in the pre- and post-test sessions, ($p < 0.01$) while no such difference occurred with B. Age was not significant either as a main term or in any interaction.

Thus the experiment showed that young children transfer information over a transformation when the second display is a neutral one, but not when the second display normally produces a judgment opposite to the one produced by the first display.

Discussion

These results suggest that children as young as three do understand and can use the principle of invariance. When changes in response do occur after a transformation they can be explained in terms of a resolution of a conflict between incompatible hypotheses. Children have consistent hypotheses about B as well as about A, although the former are incorrect hypotheses. The very fact that they can have incorrect hypotheses must mean that they have no systematic way of working out what is a correct and what an incorrect hypothesis. Thus the A-B transformation faces them with a conflict between incompatible hypotheses and no systematic way of resolving it. Their solution is to adopt the more recent hypothesis.

The main feature of this analysis is that as far as the young children are concerned their B judgment is as good as their A judgment, and so there is no reason why they should maintain the latter at the expense of the former in an A-B transformation. This analysis produces two predictions. The first is that transfer from B to C in a B-C transformation should be as effective as in an A-C trial, but in the reverse direction. After a B-C transformation children should perform at a below chance level with the C display. The second prediction is that children faced with a transformation from B to A should switch hypotheses exactly as they did in the A-B trials, though again in the opposite direction. Both predictions are that transformations from B should have exactly the opposite effect from transformations from A. The second experiment therefore investigated performance with B-A and B-C transformations.

Experiment II (B-A, B-C)

Method

Subjects. There were four age groups with sixteen *S*s to a group. The mean ages of the four groups were:

(*a*) 3 yrs 9 mths (range 3 yrs 7 mths–3 yrs 11 mths);
(*b*) 4 yrs 6 mths (range 4 yrs 2 mths–4 yrs 9 mths);
(*c*) 5 yrs 7 mths (range 5 yrs 6 mths–5 yrs 9 mths);
(*d*) 6 yrs 6 mths (range 6 yrs 4 mths–6 yrs 8 mths).

These were taken from two Nursery schools and one Infant school in Oxford. The *S*s had not been tested previously.

Displays. The same three displays were used as in the first experiment.

Procedure. There were three sessions as in the first experiment, a pre-test, an experimental and a post-test session, in that order. The procedure in the baseline pre- and post-tests was the same in the first experiment. The procedure in the experimental session was also the same as in the first experiment, except that it involved two different transformations. The children were given eight B-A transformations (four trials with A_1 and four with A_2, and eight B-C transformations.

Results

The mean scores for the three sessions are presented in Table 2. As in the first experiment the baseline pre- and post-tests established that performance on A presented on its own is above chance, on B well below chance, and on C at chance level.

Table 2 *Second experiment (*B – A: B – C*): mean correct on* A, B *and* C *displays*

			Pre-test	*Exp. Session*	*Post-test*				*Pre-test*	*Exp. Session*	*Post-test*
Three-year group	A	*X*	7.19	6.81	7.25	Five-year group	A	*X*	7.38	7.44	7.15
		SD	0.90	1.00	0.90			*SD*	0.93	1.17	0.56
	B	*X*	1.19		0.81		B	*X*	0.38		1.00
		SD	1.13		1.07			*SD*	0.48		1.06
	C	*X*	3.38	1.56	4.38		C	*X*	3.44	0.63	4.00
		SD	1.41	1.41	1.05			*SD*	1.77	0.86	1.84
Four-year group	A	*X*	6.69	6.63	6.93	Six-year group	A	*X*	7.31	7.13	7.56
		SD	0.81	1.22	0.90			*SD*	1.04	1.45	0.70
	B	*X*	0.63		0.63		B	*X*	0.88		0.56
		SD	0.63		0.52			*SD*	1.05		0.79
	C	*X*	4.31	1.19	3.13		C	*X*	3.13	0.81	4.19
		SD	1.10	0.73	0.99			*SD*	1.69	1.07	1.55

(Exp. session scores for A and C are after transformation from B.)

In the experimental session performance on C in the B-C trials fell well below chance level. Since the transformation was from B, on which baseline performance is below chance, it must be assumed that the deterioration on C in the experimental session is due to a transfer of an hypothesis across a transformation. This time the hypothesis which is transferred is an incorrect one.

Performance on A did not differ between sessions. It was always well above chance, and remained so as much in the experimental session as in the other two sessions. Thus young children again transferred a quantity judgment over a transformation to a display which would not on its own have produced this hypothesis (B-C), and again failed to transfer when the second display produced an hypothesis contrary to the pre-transformation hypothesis (B-A).

The scores were analysed in a three-way analysis of variance, in which the main terms were Ages, Displays, and Sessions. The main result of the analysis was a significant Displays × Sessions interaction (*F* 53.093; *df* 2,120; $p < 0.001$). Tukey tests showed that the reason for this interaction was that whereas performance on C was worse in the experimental session than in the pre-test ($p < 0.01$) and the post-test ($p < 0.01$) sessions, no such difference between sessions occurred with the A displays.

Thus this experiment demonstrated that young children transfer information from a B display in much the same way as they do from an A display. Once again they were able to transfer a quantity judgment over a transformation to a neutral display, which on its own would not have produced this judgment consistently and once again they did not transfer when the judgment ordinarily produced by the second display was in direct conflict with that ordinarily produced by the first.

Discussion

The two experiments produced two highly consistent results. The A-C and B-C trials showed that very young children can maintain a quantity judgment and transfer it to a display which on its own would not have produced this judgment. However, the A-B and B-A trials showed that this transfer breaks down when the responses normally produced by the two displays conflict.

The first result suggests that young children do have an effective grasp of the invariance principle and can use it. This is an important conclusion, since this is the first strong evidence for the understanding of invariance at such an early age. Other attempts to demonstrate the invariance principle in very young children (Bever, Mehler and Epstein, 1968; Mehler and Bever, 1967) have been shown to be fraught with methodological errors (Rothenberg and Courtney, 1968; Siegel and Goldstein, 1969). The result supports the hypothesis conflict analysis of performance in conservation tasks.

This analysis can also deal quite happily with the second result, since the suggestion is that the A and B displays produce a conflict which the young children have no way of resolving. Instead they arbitrarily select the more recent judgment, which is the one based on the B display in the A-B trials and on A in the B-A trials.

If this analysis is correct children must eventually 'conserve' not because they discover invariance (which they know about already) but because they learn that some cues are good and others bad. They learn, for example, that the one-to-one correspondence cue is consistent and highly reliable, while the length cue is inconsistent and highly unreliable, at any rate in number comparisons. It follows that young children who have been trained to distinguish which of these two hypotheses is the better should no longer make errors in A-B transformation trials. They should realize that the A hypothesis is the superior one, and that it should be maintained at the expense of the B hypothesis.

It is very easy to devise ways of training young children to distinguish the relative worth of the two cues simply by demonstrating that the one-to-one correspondence cue is entirely reliable over transformations, while the length cue is entirely unreliable. This can be done by showing children two series of displays, as in Figure 2. In one series (A) children will tend to use the one-to-one correspondence cue, in the other (B) the length cue. They can then be shown the two series separately in such a way that in one set of trials A_1 is transformed into A_2 which is then transformed into A_3 and so on, while in the other B_1 is transformed into B_2 which is then transformed into B_3, etc. Thus in one series A-type displays are transformed into each other, while in the other B-type displays are transformed into each other. Note that in the A series the child's judgment will necessarily be consistent if he uses the one-to-one correspondence cue. In the example, the black row

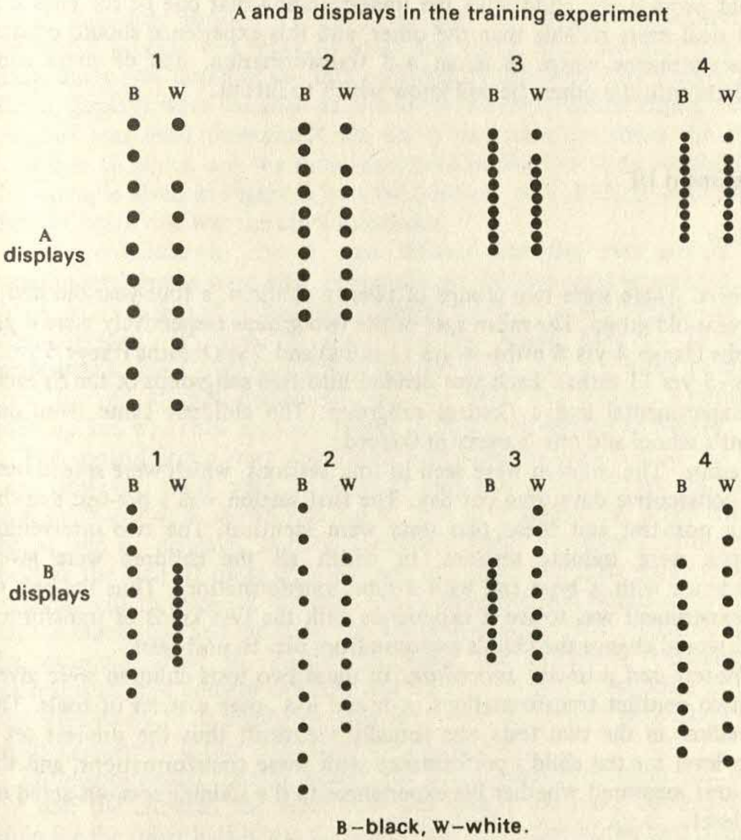


Fig. 2 Displays used in the third experiment

will always be judged as more numerous, as indeed it is. On the other hand, in the B series the black row will be judged to be more numerous only in displays B_1 and B_2 if the length cue is used: in the other two if the child will judge, quite erroneously, that the white row is the more numerous one.

If the hypothesis conflict analysis is correct experience with both series should provide the child with the understanding that one of the cues is a great deal more reliable than the other, and this experience should change his performance when, as in an A-B transformation, one of these cues conflicts with the other. He will know which to favour.

Experiment III

Method

Subjects. There were two groups of twenty children, a four-year-old and a five-year-old group. The mean ages of the two groups respectively were 4 yrs 9 mths (range 4 yrs 6 mths–4 yrs 11 mths) and 5 yrs 8 mths (range 5 yrs 3 mths–5 yrs 11 mths). Each was divided into two subgroups of ten *S*s each, an Experimental and a Control subgroup. The children came from one Infant's school and one Nursery in Oxford.

Procedure. The children were seen in four sessions, which were spread over two consecutive days, two per day. The first session was a pre-test and the last a post-test and these two tests were identical. The two intervening sessions were training sessions, in which all the children were given experience with A-type and with B-type transformations. Thus the aim of the experiment was to see if experience with the two kinds of transformations would change the child's response from pre- to post-tests.

Pre-test and post-test procedure. In these two tests children were given the two conflict transformations, A-B and B-A, over a series of trials. The procedure in the two tests was actually identical: thus the pre-test set a basic level for the child's performance with these transformations, and the post-test measured whether his experiences in the training sessions acted on this level.

There were sixteen trials in each test session, eight of them A-B transformations and eight B-A transformations. The same displays and the same procedure applied in these trials as in the equivalent trials in the experimental sessions of Experiments I and II. The order of the trials and the position of the more numerous row were varied in the same way as in these experiments.

Training sessions procedure. Each training session contained only two trials, but each of these trials involved thirty-one consecutive transformations. In one of the trials the four A displays in Figure 2 were transformed into each other, while in the other the B displays were transformed into each other. All the children were treated identically in the A trial. However, the experimental and control subjects received different versions of the B trial. The experimental children were given transformations which were likely to lead to conflict, while the control children were only given transformation which led to consistent responding.

The A *trial* started with one of the A displays in Figure 2. The child was asked which row was more numerous, and then the counters were transformed into another A display: the question was asked again, and then the whole process was repeated again and again, until the child had been given thirty-two displays, i.e. until each display had been shown eight times. The A displays were designed as one-to-one correspondence displays, and if this cue was used throughout the thirty-one transformations the child's answer as to which was the more numerous would have to be consistent. In the example given in Figure 2 he would answer with all thirty-two displays that the black row was the more numerous.

The procedure in the B *trial* differed for the two groups. The experimental group were also shown all four displays eight times each, as in the A *trial*. It was expected that children would use the length cue with these displays. If they did their response would have to be highly inconsistent, since they would say that the black row was more numerous given B_1 and B_2 and, incorrectly, that the white row was more numerous given B_3 and B_4. Their response would change from display to display.

The control group were also shown thirty-two displays in much the same way. However, these were made up of only two displays. Five children in each subgroup were shown just the B_1 and B_2 displays, the other five just the B_3 and B_4 displays. In both cases each display was shown sixteen times through the trial. If these children used the length cue their response would be consistent throughout the trial. The former children would always point out the black row, and the latter the white row, as the more numerous (with the example given in Figure 2). Thus the difference between the experimental and control subjects was that the former's responses in the B trial were likely to be highly inconsistent, while the latter's were likely to be highly consistent, if they both used the length cue.

Throughout each training session the colours used were black and white. For half the children the black row was always the more numerous one, while for the other half it was the white row. The order of the two trials was systematically varied between *S*s and so was the order of displays within each trial. The procedure during the two training sessions was identical.

Results

Training sessions. All the children were highly consistent in both A trials. They all made thirty-two responses in each of these trials, and in fact 96 per cent of these were correct in the four-year-old group and 98 per cent in the five-year-old group. They were less consistent in the B trial, but here there were large differences between experimental and control groups. The control groups used the length cue fairly consistently. In the four-year-old group 92 per cent of responses were to the longer row in the first training session and 95 per cent in the second. The equivalent scores for the five-year-olds were 94 per cent and 90 per cent. However, the length cue tended to be used less by the experimental subjects and their use of the cue declined from first to second training session. The four-year-old experimental subjects chose the longer row with 84 per cent of the displays in the first training session and with 75 per cent of them in the second training session. The five-year-olds chose it only 75 per cent of the time in the first and 68 per cent in the second training session. Thus there was evidence that both age groups were able to register a difference between A and B type displays when the latter produced conflict.

Pre- and post-tests. The mean scores for these sessions are presented in Table 3. This shows that both age groups responded to the A-B and B-A transformations in the pre-test in much the same way as the other children had in the first two experiments. Again performance in the A-B trials was well below, and in the B-A trials well above, chance.

Table 3 *Training experiments: mean number correct post transformation scores*

			Experimental subjects				*Control subjects*	
			Before training	*After training*			*Before training*	*After training*
	B–A Trials	*X*	7.4	7.6	B–A Trials	*X*	7.2	7.3
		SD	0.80	0.66		*SD*	0.98	1.00
Four-year group								
A–B	–AB Trials	*X*	1.7	4.6	A–B Trials	*X*	0.8	0.9
		SD	1.10	3.04		*SD*	1.12	1.22
	B–A Trials	*X*	7.4	6.9	B–A Trials	*X*	7.2	7.9
		SD	0.80	1.00		*SD*	0.75	0.30
Five-year group								
	A–B Trials	*X*	1.1	4.3	A–B Trials	*X*	1.2	1.7
		SD	1.22	1.22		*SD*	1.20	1.19

However, in the post-test, there was an improvement in the A-B trials, and in both age groups it was only the experimental children who showed this improvement. A four way analysis of variance, with Ages, Tests (Pre *vs* Post) Groups (Experimental *vs* Control) and Trials (A-B and B-A) as the main terms produced a Tests × Groups × Trials interactions (F 15.885, df 1.36, $p < 0.001$), which subsequent Tukey tests showed to be due to the fact that the Experimental *S*s' performance on the AB trials was better in the Post-test than in the Pre-test ($p < 0.05$).

Not all the experimental subjects showed this particular improvement. In the younger group three of the experimental subjects were as bad in the post-test A-B trials as they were in the pre-test: one improved by one point, and this was probably a chance result. The remainder improved by four or more points, and it can be inferred that these children had benefited from their experiences in the A and B trials of the training sessions. Much the same pattern occurred in the older group. Three of the experimental *S*s showed no improvement from pre- to post-test, two improved by three points and the remainder by four or more points.

The control *S*s of both age groups showed no such improvement. The greatest improvement among the control *S*s of the younger group was three points and this was shown by only one *S*. In the older group the greatest improvement was again three points shown by two *S*s. Most of the control *S*s produced highly similar scores in both Tests.

Summary of results. The main result was that some children who initially perform very badly after an A-B transformation perform much better when they have been taught that the one-to-one correspondence cue used with A is much more reliable than the length cue used with B. After such training they appear largely to abandon the length cue. When the one-to-one correspondence cues are pitted against length cues the children appear to trust and maintain their judgments based on the former cue, whatever the order of the displays.

Discussion

The major results of these experiments were (1) that children as young as three can maintain a quantity judgment over a transformation, but (2) that they do not do so when the response characteristically produced by the second display is opposite to that produced by the first, and (3) that training that one cue is more reliable than the other enables young children to maintain a judgment based on the more reliable cue at the expense of using the less reliable one.

These three results can be explained by the hypothesis conflict theory, and it is difficult to see how all three can be explained in any other way. According to this theory young children understand the principle of invariance, and use it unless they are faced with a conflict between incompatible hypotheses. This leaves them in an insoluble quandary as to whether they should maintain their first hypothesis at the expense of the second or instead abandon the first in favour of the second. The theory accounts for the first of the three results, as Piaget's theory would not. It also accounts for the second result since here as in the traditional conservation experiment, the judgments, separately produced by the two displays involved in each transformation, are in conflict. The hypothesis conflict theory argues that since children at this age form incorrect judgments as well as correct ones, they cannot properly distinguish between a correctly based and an incorrectly based hypothesis, and are at a loss to know how to resolve the conflict between hypotheses rationally. Instead they adopt the more recent hypothesis whichever that is.

There is an alternative explanation of the first two results, which is that young children understand invariance in situations involving no conflict, but that in conflict situations, where they do change their responses over a transformation, they genuinely think that the quantities have undergone a change. This is a watered down version of Piaget's original theory, although it approximates very closely to some of his more recent statements, which intimate that young children may have a basic understanding of invariance in some situations (Piaget, 1968). However, this alternative explanation cannot deal with the third main result. There is no reason why children should learn anything from a demonstration that a judgment based on the use of the length cue is highly unreliable if, when they alter their judgments over a transformation, they genuinely think that the quantity has changed. If they do think this, then inconsistency will be no problem, and therefore there is no reason why they should avoid using the length cue after such training. But the third experiment demonstrates that children who are shown that the one-to-one correspondence cue is reliable, and the length cue is not, will begin to favour the former against the latter with some effect.

The advantage of these experiments is that they have for the first time made a clear distinction between the formation of quantity judgments, and their maintenance over perceptual transformations. It seems that young children are perfectly able to maintain a quantity judgment, and therefore understand invariance. Where they fail, and where they must eventually improve, is in the initial formation of these hypotheses. Thus the important development is not as has generally been assumed in the aquisition of the invariance principle, but in learning what constitutes a correct and what an incorrect basis for a judgment about quantity.

These conclusions have both practical and theoretical significance. The practical implications are that the emphasis should be shifted to training children how to form judgments properly instead of teaching them invariance. In fact, experimental attempts to train invariance have met with remarkably little success (Smedslund, 1961; Wallach, 1969). This could very well be because the experimenters were trying to teach children the wrong thing, to teach them, in fact, something which they knew already.

The theoretical implications are also important. It seems that young children have a basic understanding of the invariance principle, but are unable to put it into practice completely effectively. There is a gap between what they understand and what they can do, but this gap has nothing to do with the nature of their grasp of invariance. It is due to the nature of the judgments which they have to maintain, and not to the way in which they try to maintain them. The gap can readily be described, in current terms, as a gap between competence and performance. It seems that young children's performance with some invariance problems lies far behind their competence. It is their developmental task between the ages of three and eight years not to acquire the competence but to bring their performance more in line with their competence. This is a task in which, as the third experiment shows, they might quite easily be helped.

References

BEVER, T. G., MEHLER, J. and EPSTEIN, J. (1968) 'What children do in spite of what they know' *Science, 162,* 921–4.

BINET, A. (1890) 'La perception des longueurs et des nombres chez l'enfant', *Revue Philosophique, 30,* 68–81.

MEHLER, J. and BEVER, T. G. (1967) Cognitive capacity of young children. *Science, 158,* 141–2.

PIAGET, J. (1968) Quantification, Conservation and Nativism. *Science, 162,* 976–9.

PIAGET, J. and SZEMINSKA, A. (1941) *La genese du nombre chez l'enfant.* Neuchatel: Ed. Delachaux et Niestle.

SIEGEL, L. F. and GOLDSTEIN, A. G. (1969) 'Conservation of number in young children: recency vs. relational response strategies', *Dev. Psych., 1,* 128–30.

SIGEL, I. E. and HOOPER, F. H. (1968) *Logical Thinking in Children.* New York: Holt, Rhinehart and Winston.

SMEDSLUND, J. (1961) The acquisition of conservation of substance and weight in children. *Scandinav. J. Psychol., 2,* 11–20.

WALLACH, L. (1969) 'On the bases of conservation.' In Elkind, D. and Flavell, J., *Studies in Cognitive Development.* Oxford: Oxford University Press.

WOODWORTH, R. S. and SCHLOSBERG, H. (1954) *Experimental Psychology.* London: Methuen.

Section Five
Cognitive styles

Introduction

Probably no two people approach a given learning situation in exactly the same way, whether the situation is studying a research article, learning to type or understanding a mathematical problem. Such differences are most often called cognitive styles. In this section our interest is focused on whether these differences are substantial enough to be of educational importance and, if so, whether they can be systematized in a useful manner.

The educational importance of cognitive styles derives from the difference it makes to the learner when he is taught in a way that suits his cognitive style, instead of in a way unsuited to it. Effective learning may be vastly different in the two cases. It is unlikely that it will ever be practicable (or necessary) to cater for every individual's learning idiosyncrasies, but it *is* feasible to cater for a small number of broad categories or tendencies. The three articles in this section explore these issues in very different ways.

The first, by Pask and Scott, describes experiments attempting to distinguish the details of individual learning strategies. The learners have some factual material to master and are free to set about this in a variety of ways—the free learning situation. Whichever approach they prefer, they can only proceed by asking for information, so that each learner asks a sequence of questions which can then be analysed. In the Pask and Scott study there are two broad categories of learner, or types of individual competence: the serialist and the holist. The serialist proceeds by small steps, trying to get each point absolutely clear before moving on to the next. The holist, in contrast, proceeds much more broadly, tackling several ideas at once. The difference between the two strategies is rather like that between studying a book page by page, considering each new idea until it is understood, and the very different approach of reading on fairly steadily, in the expectation that the material will eventually fall into place.

Having identified two broad classes of learning strategy, Pask and Scott go on to discuss teaching strategies based upon these. They present evidence that mismatched students (serialists confronted with teaching materials designed for holists, and vice versa) learn much less effectively than matched students, at least within the subject matter they used. The educational implications of this result would be considerable if it could be shown to extend to other content areas as well. The authors are continuing their researches along those lines.

The second and third articles in this section concern cognitive style dimensions. The interest focuses, not on detailed strategy differences, but upon more general characteristics.

Witkin summarizes a considerable amount of research into the degree to which different people are influenced by the context in which they see something, or in which they encounter a problem. He sketches the historical development of this research from its beginning in the field of perception, when the dimension was labelled field-dependence/independence, to its more general application. The relevance of this cognitive style for education is explored, particularly in connection with its relationship to the IQ. This is not a simple relationship, because the IQ involves a variety of skills, only some of which correspond closely with the skills required to see things independently of their context.

The article by Kagan, Pearson and Welch is concerned with the cognitive style dimension known as conceptual reflection/impulsivity. The interest here is in the degree to which an individual thinks before proffering an answer to a question, in how much he reflects before responding. At one extreme of the dimension the impulsive person tends to respond quickly, and often inaccurately, while at the other end the reflective individual ponders until he is virtually sure he is right before making any response.

The article discusses the significance of this cognitive style when children are asked to perform certain inductive reasoning tasks. The authors demonstrate that, within the context of these tasks, relatively impulsive children meet with much less success than do reflective children. As the skills involved in making inferences are important in many learning situations–indeed, discovery methods depend on such skills–it is important to help children to acquire them.

Thus all three articles have considerable educational relevance, albeit of very different kinds. What they all have in common is the aim of identifying and catering for significant differences in cognitive styles.

Ann Floyd

20 Learning strategies and individual competence

G. Pask and B. C. E. Scott

1. Introduction

A student learns in a way that depends upon his individual cognitive style. It is less commonly recognized that differences of style have great educational importance because students fare very differently according to whether the teaching materials are or are not adapted to suit their idiosyncracies. Henceforward, a student's style or mental character is referred to as his cognitive *competence* which, as we shall see later, is a measurable individual property.

Likewise, teaching materials can be structured, via the tutorial strategy they embody, to fit one of two or three types of commonly observed competence. If the teaching materials are structured to suit the individual's competence (a 'matched' condition), then learning is very much more effective than it is if the converse applies (a 'mismatched' condition).

The experiments reported in this paper were concerned with establishing differences in competence, with setting up matched and mismatched conditions and comparing their influence upon learning, both at a molar and molecular level. In addition, a quantifiable 'teachback' testing technique was introduced which, superficially at least, appears more defensible than multiple choice (or other readily scored) examination methods.

1.1 *Practical application of the results*

The practical consequences of the (generally affirmative) results to be reported are as follows: (*a*) education could be rendered much more effective if students were assigned properly matched tutorial materials. As it stands, most everyday materials, lectures, texts, teaching programmes, have a strong bias towards one or another type of student. Since they are presented indiscriminately to a mixed class any individual is likely to receive mismatched instruction; (*b*) the 'proper' design principles can be clearly specified and are neither hard to instrument nor counter-intuitive; (*c*) two or three types of individual difference in competence can be quite easily detected, as a result of which the 'matching' requirement is entirely realistic; and (*d*) conventional multiple choice examinations could be profitable replaced by teachback testing (which would, incidentally, provide *one* source of the data needed to establish a student's competence).

Source: *International Journal of Man-Machine Studies,* 1972, *4,* 217-39, 242-53 (abridged).

1.2 *Outline of procedures*

Quite simple tests, which may either be manually or automatically administered, are shown to distinguish between two major categories of mental competence which reflect an individual's cognitive style and lead him, in free learning conditions, to prefer a certain type of strategy, i.e. to direct his attention in a specific way to different parts of a learning task, to ask specific sorts of questions, to assimilate material by specific types of 'self explanation' and to pose specific sorts of hypotheses.

The two major categories of individual are called *serialists* and *holists*. This distinction was first made in the context of a relation learning task (Pask, 1969a; Pask and Scott, 1971) but it has been found equally germane in connection with intellectual tasks.

Serialists learn, remember and recapitulate a body of information in terms of string-like cognitive structures where items are related by simple data links: formally, by 'low order relations'. Since serialists habitually assimilate lengthy sequences of data, they are intolerant of irrelevant information unless, as individuals, they are equipped with an unusually large memory capacity. Holists, on the other hand, learn, remember and recapitulate as a whole: formally, in terms of 'high order relations'.

There are two subcategories of holist called *irredundant holists* and *redundant holists*. Students of both types image an entire system of facts or principles. Though an irredundant holist's image is rightly interconnected, it contains only relevant and essential constituents. In contrast, redundant holists entertain images that contain *logically* irrelevant or overspecific material, commonly derived from data used to 'enrich' the curriculum, and these students embed the salient facts and principles in a network of redundant items. Though *logically* irrelevant, the items in question are of great psychological importance to a 'redundant holist', since he uses them to access, retain and manipulate whatever he was originally required to learn.

The categories of competence (serial, holistic) are quite commonsensical and would be evident to many experienced teachers. Our innovation is twofold:

1. Although the categories could be empirically adduced, it is also possible to predict their existence on the basis of a theory of learning and teaching; the prediction is logically exact and unambiguous. The theory is described in Pask, 1965, 1968, 1969b,c, 1970a,b, 1971.
2. Due to the clearcut distinction existing at a theoretical level, it is possible to specify simple testing methods for determining an individual's type of competence and (as later) to show that these methods are effective.

For experimental purposes, an individual's competence was determined under 'free learning' or 'browsing' conditions by a questioning and goal selecting procedure that led him to externalize his mental process as an observable stretch of behaviour.

The free learning technique (Section 2) has the following general properties. Prior to free learning the subject is introduced to the task goal (learning a zoological taxonomy) and reads through an introductory and irredundant description of the system. Now the taxonomic system and its universe of discourse can be redundantly described by five categories of statements which are inscribed on cards placed in separate piles according to their category. The student learns by 'questioning', i.e. by turning up cards and (if he wishes) generating new 'cards' via note-taking. He is given a limited period in which to work but he is free to ask any question provided that he states a *reason* for doing so. The 'questions' asked, the reasons for asking them and the notes taken (and counted as 'fresh cards') are all recorded. As anticipated from preliminary studies, the reasons could all be summarized under one of the following headings:

> (i) to 'search a co-ordinate of the message space', i.e. to pass through a pack of cards as though thumbing over the pages of one chapter in a book; (ii) to 'search for a datum regarding a particular object, specimen, fact, etc., named independently of the card-category co-ordinates'; (iii) to 'test an hypothesis about a simple predicate' (for example, that a specimen in Class X has more than two legs); (iv) to 'test an hypothesis about a complex predicate' (for example, that Class X differs from Class Y in respect of leg number *and* mating behaviour *or* habitat).

The serialist differs from the holist chiefly in terms of the complexity (order) of predicates involved in the hypotheses he tests. A redundant holist differs from an irredundant holist in terms of his question distribution amongst the card-category co-ordinates. Success and failure in a subsequent test are correlated with the ratio of the number of hypotheses testing questions to the number of searching questions uttered during free learning (a better correlation seems to exist between success and the *logical* information content of the hypothesis testing questions.)

The tenure of the competence type assigned on the basis of a student's free learning characteristics was checked by a simple content analysis of the dialogue during an interview-like or tutorial procedure in which each student 'taught back' to the experimenter whatever he had learned.

The analysis extracts (amongst other things) the proportion of inferential and hypothetical statements uttered during teachback and the value of this quantity discriminates between serialists and holists. In all cases, this molecular teachback data yielded exactly the same competence assignment as the molecular free learning data.

It is also true that either method of competence assignment (free learning, teachback) may be 'automated' for use in a CAI [computer-aided instruction] system. Experimenter administration, though convenient and entirely practicable, is not mandatory. For, once the data base is fixed, the administration follows an algorithm.

Finally, there is some evidence that a competence type can be reliably assigned on the basis of a simple content analysis of essays or other written materials that are freely composed by the student in question.

Thus 'competence determination' might be practically accomplished by any one of the following methods.

(*a*) Questioning and selecting during free learning–(human adminstration).
(*b*) Questioning and selecting during free learning–(machine administration).
(*c*) Content analysis of essays (for example, using the categories outlined by Schneidmann (1966)).
(*d*) Tutorial (teachback) procedure–(human administration).
(*e*) Tutorial (teachback) procedure–(machine administration).

Any piece of educational material, a textbook, a course module or a teaching programme, is based upon a teaching strategy. The teaching strategy is analagous to the learning strategy used by a student to direct his attention during free learning except that it is imposed by the teacher or course designer rather than generated by the student himself.

Teaching strategies can be slanted to suit an individual with a given sort of competence. For example, a teaching strategy can be designed with a serialist in mind, or a holist.

The theory prescribes quite accurately, what *sort* of teaching strategy *will* suit a given type of individual. When speaking of 'serialist materials' and 'holist materials' in the sequel we mean materials structured according to these principles.

Since human beings are highly adaptable it may be possible for an individual with *any* sort of competence to learn, in the end, according to any teaching strategy. But the experiments show, very clearly indeed, that the rate, quality and durability of learning is crucially dependent upon whether or not the teaching strategy is of a sort that suits the individual. The 'matched' and 'mismatched' conditions were compared as shown in Figure 1.

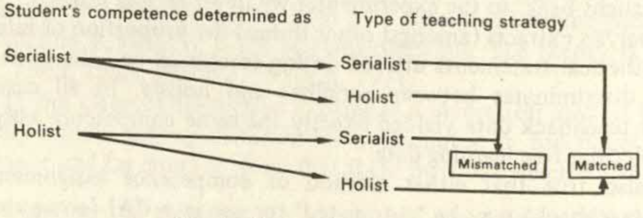


Fig. 1 Comparisons made

Whereas the matched students learned effectively (near perfectly) the mismatched students did not do so (and retained hardly any ungarbled

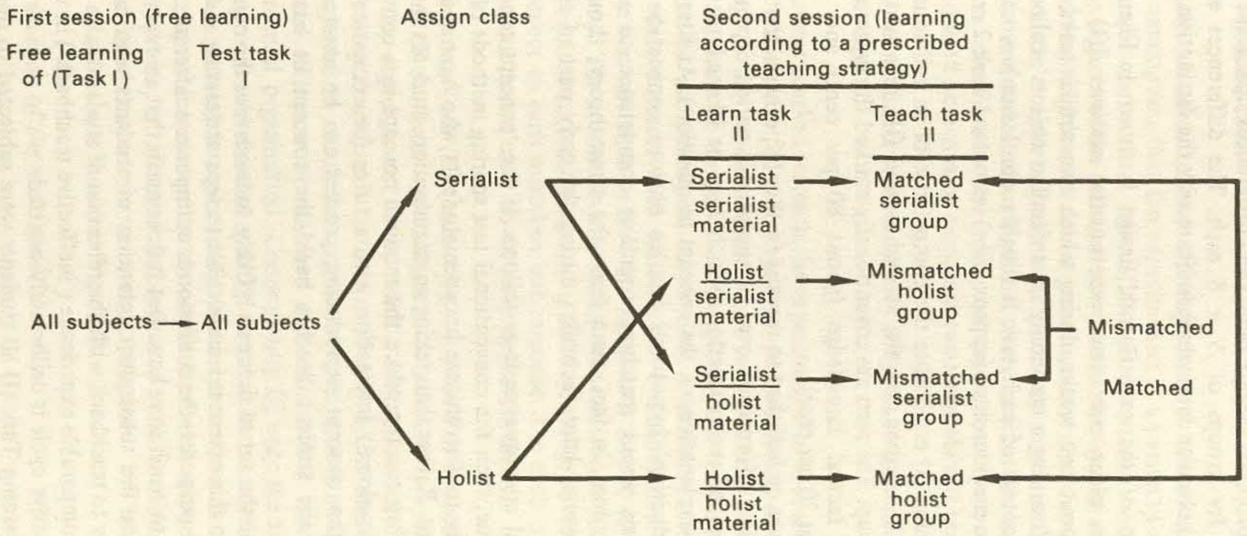

First session (free learning)
Free learning of (Task I)
Test task I
All subjects
All subjects
Assign class
Serialist
Holist
Second session (learning according to a prescribed teaching strategy)
Learn task II
Teach task II
Serialist serialist material
Holist serialist material
Serialist holist material
Holist holist material
Matched serialist group
Mismatched holist group
Mismatched serialist group
Matched holist group
Mismatched
Matched

Fig. 2 Outline design

knowledge). This assertion is based on molar data; the test results of section 2 where the appropriate comparisons yield differences significant at the 0.1 per cent level for groups of $N = 8$ each. The differences evident on examining the molecular data, which indicate why the distinction exists, are even greater.

The skeleton of the experimental design is shown in Figure 2. The treatment takes place over two experimental sessions [(I) and (II)] separated by about two weeks, during which two similar tasks were used (both of them 'learning a taxonomy of unfamiliar objects', called 'Martian Fauna'). The content of each task is roughly one lesson's worth; Task 1 corresponds to a medium-sized (approx. ½hr) lesson and Task 2 corresponds to a long (approx. 1 hr) lesson.

The first session (I) establishes the competence of an individual student; as a result he is assigned, on the second session (II) to a matched or a mismatched group. The *tests* are coventionally-marked thirty-item questionnaires tapping factual knowledge (about 80 per cent) and ability to generalize (about 20 per cent).

The design was embellished to test the further hypothesis, that an ability to retain complex information over an appreciable period depends upon a cognitive process of reconstruction (which may be likened to the reproduction of self-explanation of the concept in question). At first sight, the 'concept reproduction' hypothesis has no obvious connection with the initial hypothesis about matching cognitive competence to a teaching strategy. But it does, in fact, stem from the same theory, though it was developed somewhat later (actually, during the early part of the experimental work).

The practical consequence is as follows. If the 'concept reproduction' hypothesis is true, then the conventional test scoring methods of examination are unlikely to discriminate between students who *have* and have *not* learned a concept. Rather than taking an examination, students should show evidence that they can reproduce the required concept in a conversational (or quasi-conversational) interaction with a tutor (or computer program) during which the concept reproducing process can be externalized for observation as the student 'teaches back' the concept he has allegedly learned.

It is true that the act of doing so is likely to teach *him* the concept in its own right and to this extent it is impossible to separate testing and teaching. For educational purposes, the confusion is unimportant; after all, the object of education is to *teach* somehow. But it does mean that an experiment can only discriminate the subsequent retention of students who have demonstrated an ability to teachback with the retention of students who have been subjected to a comparable experience ('ineffective teachback') in which part of the reconstructive cycle is deliberately omitted.

After free learning Task (I) all students were subjected to one or other version (*a*) or (*b*) of a teachback procedure.

Half of them were randomly assigned to a real teachback (unqualified teachback) defined thus:

(*a*) the student is asked to teach the experimenter how to use the taxonomy on the assumption that the experimenter is a person of the same mental make-up as the student himself but is omniscient with respect to all *relevant* aspects of the task (i.e. *to all items that appear in the test and all items, derivable from the irredundant introduction that might appear in the test*). On completing the teachback any mistakes or omissions are corrected and the student is asked to reconstruct his teachback so that they are eliminated. The cycle of *correction* and *reconstruction* is repeated until no errors or omissions remain in the teachback.

Half of the students were randomly assigned to an ineffective teachback defined thus:

(*b*) the student is asked to teach the experimenter how to use the taxonomy on the assumption that the experimenter is a person of the same mental make-up as the student himself but is *omniscient* with respect to all *relevant* aspects of the task. On completing the teachback any mistakes or omissions are noted and the student is furnished with knowledge of results that corrects each mistake (or omission) *and* states *why* the response was mistaken or how an omission is related to the body of the taxonomy. The student is now *tested* over omitted or mistaken items (he is *not* required to teachback again) and knowledge of results is furnished with respect to each mistaken test response. The cycle of *correction* and *testing* is continued until no errors or omissions remain.

After learning Task (II) (taught by a matched or a mismatched teaching programme) all students were subjected to a further teachback routine. In this case, the object is not to establish a distinction between 'Teachback' and 'Ineffective Teachback'. Rather, as already noted, molecular teachback data is employed to check the competence assignment and, in addition, to elucidate the influence of matched and mismatched instruction upon a student's mental organization. Consequently (i) only the first cycle of the teachback process is recorded and (ii) the underlined relevance criterion of the previous account is replaced by 'the experimenter is a person of the same make-up as the student but is omniscient with respect to all relevant aspects of the task, i.e. *all items that appeared in the teaching programme'*

The modified experimental design, obtained by incorporating the teachback conditions, is shown in Figure 3.

1.3 *Generalization of the main result*

One reasonable criticism of the pilot experiment is that the initial and the

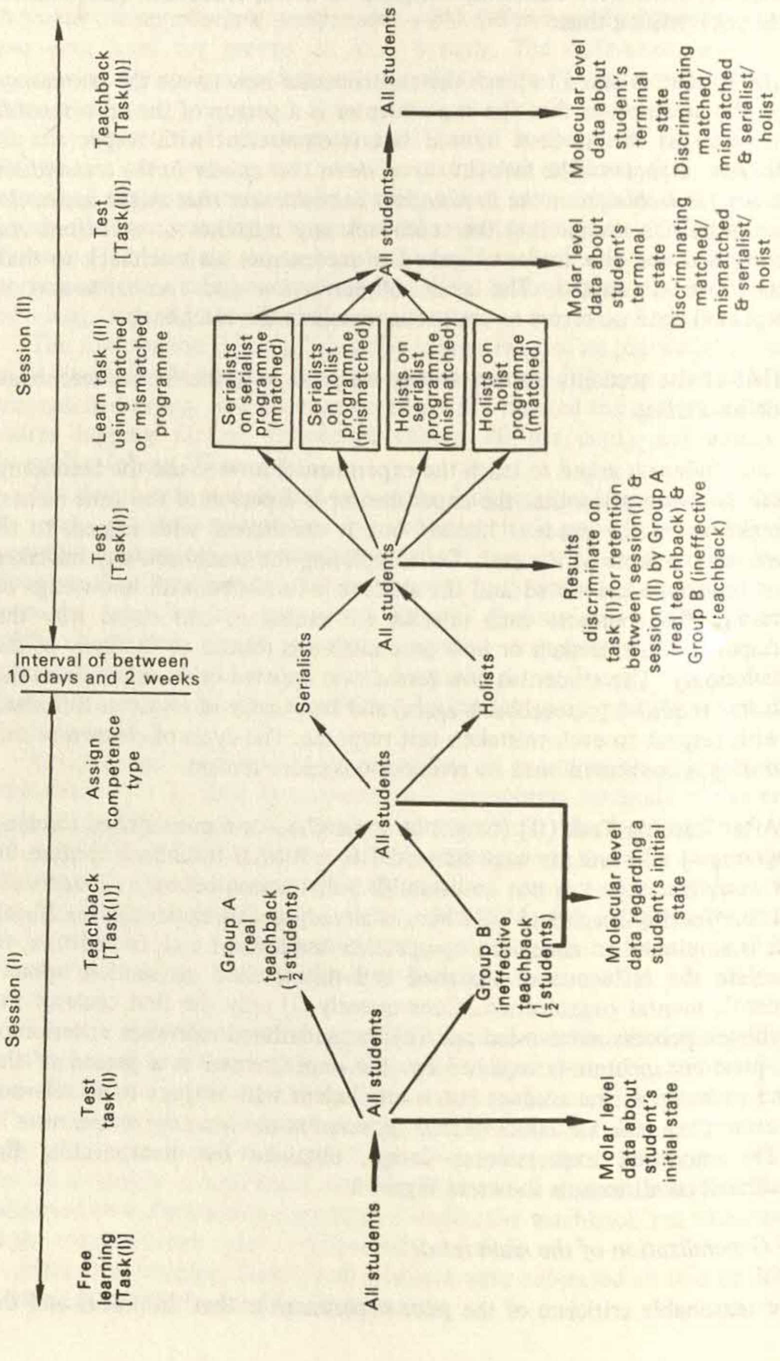

Session (I)
Interval of between 10 days and 2 weeks
Session (II)
Free learning [Task(I)]
Test task(I)
Teachback [Task(I)]
Assign competence type
Test [Task(I)]
Learn task (II) using matched or mismatched programme
Test [Task(II)]
Teachback [Task(II)]
All students
All students
Molar level data about student's initial state
Group A real teachback ($\frac{1}{2}$ students)
Group B ineffective teachback ($\frac{1}{2}$ students)
Molecular level data regarding a student's initial state
All students
Serialists
Holists
All students
Results discriminate on task (I) for retention between session(I) & session (II) by Group A (real teachback) & Group B (ineffective teachback)
Serialists on serialist programme (matched)
Serialists on holist programme (mismatched)
Holists on serialist programme (mismatched)
Holists on holist programme (matched)
All students
Molar level data about student's terminal state Discriminating matched/ mismatched & serialist/ holist
All students
Molecular level data about student's terminal state Discriminating matched/ mismatched & serialist/ holist

Fig. 3 Final design

criterial tasks are (necessarily) of the same type and thus that the match/mismatch paradigm may be restricted to just one sort of subject matter. Though valid, the criticism is unavoidable for an experiment of this magnitude.

To blunt its force, further experiments were carried out using some of the original, competence-assigned, subjects who learned a very different task under conditions of match and mismatch. These subjects had to learn and manipulate the biochemical concept of an 'operon', as proposed by Jacob and Monod (1961). A successful student must become familiar with the operation of two coupled feedback cycles, which are complicated by a double negation; in the subsystem of repressor, corepressor, and aporepressor.

Due to time pressure, only the molar data has been recorded and the result (that matched subjects fare well on a *different* type of task whereas mismatched subjects fare badly) should be taken as no more than suggestively affirmative. It is true that the gross score differences (section 2) are *signifiant* but, in the absence of molecular data obtained in the main experiment, we cannot be sure of the mental mechanism involved and thus have less confidence in the finding.

1.4 *Change of strategy*

A further criticism is that the students involved in the experiment did not change their strategy in the course of learning. The effect of strategy change is unobtrusive, partly because the learning time is relatively short and partly because the task structure is homogeneous and chosen so that it can be learned with roughly equal facility by a serialist and a holist. The *one* subject who did change his strategy was excluded from the group in order to avoid complicating the analysis and the data record.

In general, strategy change is neither uncommon nor unexpected. This aspect of the learning process has not, so far, been explored in the context of intellectual skills, but a respectable body of data is available from relation learning tasks (Pask and Lewis, 1968; Pask, 1969a, 1970a; Pask and Scott, 1969, 1971) and similar mechanisms may operate in the present field. . .

2. Experimental work

2.1 *Preliminary work and preparation of the learning tasks*

Given an outline of the experimental design in Figure 3, it was first necessary to choose an appropriate task.

Certain task requirements are mandatory.

1. Learning should occupy about 1hr for the criterial task [Task (II)] and somewhat less for the initial task [Task (I)] (i.e. a short enough time to be

experimentally tractable and a long enough time (*a*) to be realistic and (*b*) to let the student externalize a sensible stretch of mental activity).

2. The criterial and initial tasks should be similar according to any reasonable cannons.

3. It should be possible to learn according to a serialist or a holist strategy so that both kinds of strategy *can* be employed and so that a student's disposition, manifest during free learning, may be taken to indicate his competence.

In addition, it is desirable to satisfy the further conditions listed below.

4. As well as being learnable by either strategy, the task should neither be biassed in favour of a serialist or a holist teaching strategy. The intention is to compare serialist/holist students or matched/mismatched students, and not to examine the appropriateness of one approach or the other to the task in hand. Hence, an attempt was made to secure no significant difference between the mean scores, on the criterial task (II), of the serialist and the holist groups.

5. To avoid spurious individual differences, the material to be learned should not be much more familiar to one student than another.

To satisfy the requirements (1-5) an extensive series of tests were carried out using materials culled from Upton and Sampson's (1963) *Creative Analysis.* Each item used for this purpose took between five and ten minutes to learn and five separate items were examined; learning the meaning of a written passage, learning a map of the Carribean, learning a taxonomy, learning analogies and learning the operation of an abstract system. Each subject's learning process was examined by recording his recall and by informal discussion. It transpired that all of the materials chosen satisfied conditions (1), (2), (3) and (5) and taxonomy learning or, with some minor qualifications system learning were likely to satisfy condition (4).

The brief test items from *Creative Analysis* are not long enough to serve, as they stand, for experimental tasks, so four appropriate and lengthy batches of task material were constructed at this stage, and pre-tested for general suitability. The first two of them underlie Task (I) and Task (II) of the main experiment. They are two separate universes of discourse with redundantly specified descriptive properties. The objects in the universe of discourse are variously related 'Martian Fauna' called [in Task (I)] 'Clobbits' and [in Task (II)] 'Gandlemullers'. In each case a task is generated by defining a taxonomy or classification scheme, the tests in which refer to a subset of the descriptive properties (so called 'physical variables') and by requiring the student to learn (and learn to use) the taxonomy in question. These tasks are detailed in section 2.2.

The remaining two batches of material involve systems.

One is the biochemical system controlled by an 'Operon' (in the sense of Jacob and Monod, 1961) as exemplified by the regulation of lactose-glucose-metabolism. The system to be learned involves two major interactive feedback loops; the notions of inducer, repressor, corepressor and aporepressor; the (underdeveloped) notion of enzymes as specific proteins that catalyse the transformation of metabolites and the tacit but deliberately suppressed concept of the messenger RNA-Ribosome-transfer RNA chain that is responsible for synthesizing enzymes under the basic instructions written onto the DNA configuration of the structural genes.

The other batch of material concerns the human menstrual cycle and the 'underdeveloped' concept of its interruption by fertilization and pregnancy. This system involves the following basic notions: the endometrium as a growing (and later decaying) tissue controlled by oestrogen and progesterone; the ovary as a source of ova developing in their follicles via an independent growth process; the lutenization of a graffian follicle; the ovarian constituents (graffian follicle and corpus luteum) as the most relevant source of the regulating hormones (oestrogen, progesterone) and the existence of other sources (e.g. placenta, in pregnancy); the pituitary as a source of the ovary regulating hormones (follicle stimulating, lutenizing, and prolactin); the notion of feedback from oestrogen-progesterone levels to regulate the output of the pituitary hormones; the notion of overriding hypothallamic control of the feedback from oestrogen-progesterone levels to regulate the output of the pituitary and the notion of direct feedback to the hypothalamus.

With the caveats and suppressions noted, members of the target population are substantially ignorant of the systemic details to be learned and do not entertain gross misconceptions about what actually goes on. They do, of course, have a broad background knowledge (so that terms like 'DNA' or 'ovary' or 'endometrium' have meaning to them). This knowledge allows them to erect complex associational references (which is an advantage in studying the behaviour of a redundant holist) but it is of no great help to *them* in elucidating system dynamics.

Due to time limitations only one of these batches, the 'operon' system, was employed as a learning task in the present enquiry and this one formed the background of the *auxiliary* study.

2.2 *Tasks used in the main experiment*

Task (I) was to learn the 'clobbits' taxonomy. It is noteworthy that any zoological taxonomy (and, in practice, many other sorts of taxonomy also), can be learned and manipulated in various ways. For example, though a novice is *instructed* to use the classification rules in an orderly fashion, the expert unashamedly *recognizes* the member of a species by diverse, though relevant, criteria, This possibility comes about because the relations between subspecies are multiple (perhaps infinite) and can be expressed free of the taxonomy; because there are relations of symmetry, assymetry, complementation, etc., *within* the given taxonomic tree; and because there are many alternative ways of making a pertinent distinction between subspecies. For example, the subspecies might be distinguished in terms of the physical variables which generally (and, in this case certainly) are tested in the taxonomy itself; by differences in the behaviour of the animals concerned or by contextual information; for instance, regarding the habitat of the animals or the historical origins of the classifying nomenclature. We stress that this degree of richness *is* and is *known* to be embodied in our Tasks. Learning one of them is not simply an arid and abstract exercise.

In Task (I) there are 10 subspecies of 'Clobbit' with two main subtypes, 'Clobs' and 'Bits'. There are nine test boxes or branches in the classification tree (Figure 4) and seven physical characteristics are tested to make the requisite distinctions. For parts of the taxonomy, the values of *other* physical characteristics discriminate the classes equally well. The structure contains symmetries (for example, four four-legged Bits and four three-legged Bits); it also contains assymmetries (for example, the four-legged Bits are first assigned to two subspecies yielding one and three subspecies (by leg position); the three-legged Bits are assigned by leg position to two subsets each containing two subspecies. Some of the names of the animals are assigned rationally (T stands for Triped; K stands for legs at corners, but since TK is used only for three-legged creatures the suffix TK is *redundant*). Some names have no rational justification. There is background information about the behaviour and habitat of the clobbits, some of which is relevant to the distinctions made (and might be used to make them *redundant*) and some of which is irrelevant. Finally, each subspecies of Clobbit is associated with a picture of one member of it.

In Task (II) students learn a taxonomy of 'Gandlemullers' of which there are thirteen subspecies and three main subtypes. The classification tree (Figure 5) has eleven 'tests' or branch points entailing six physical characteristics. All of the general comments made about the 'Clobbits' apply equally to the 'Gandlemullers'.

2.3 *Subjects*

The sixteen subjects used in the main experiment were aged between eighteen and twenty four. None of them had taken part in the preliminary experiments (section 2.1). They were drawn from various student populations, most of them coming from Kingston College of Technology. Eight of the subjects were serialists and eight of them holists.

The balanced 8/8 sample is not altogether fortuitous and should not be taken to reflect an even distribution of serialists and holists in the population. Although the first subjects were recruited 'randomly', the remainder were selected with foreknowledge of their personality and occupation, to obtain serialists or holists as required. It is possible to infer a

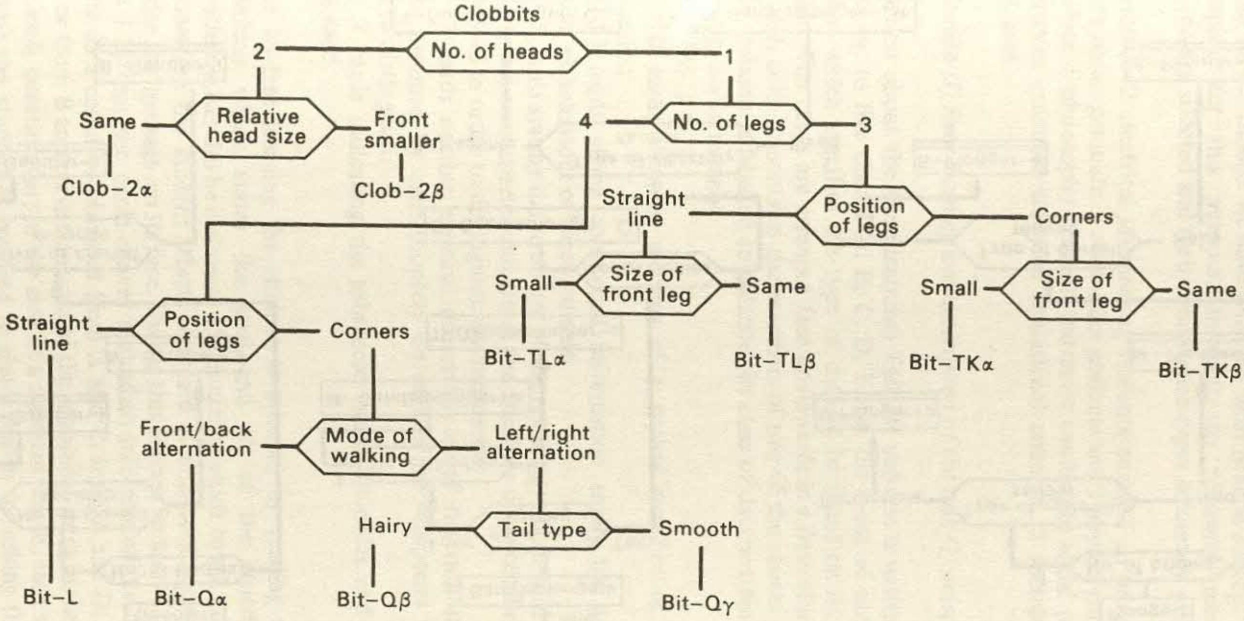


Fig. 4 Clobbit taxonomy

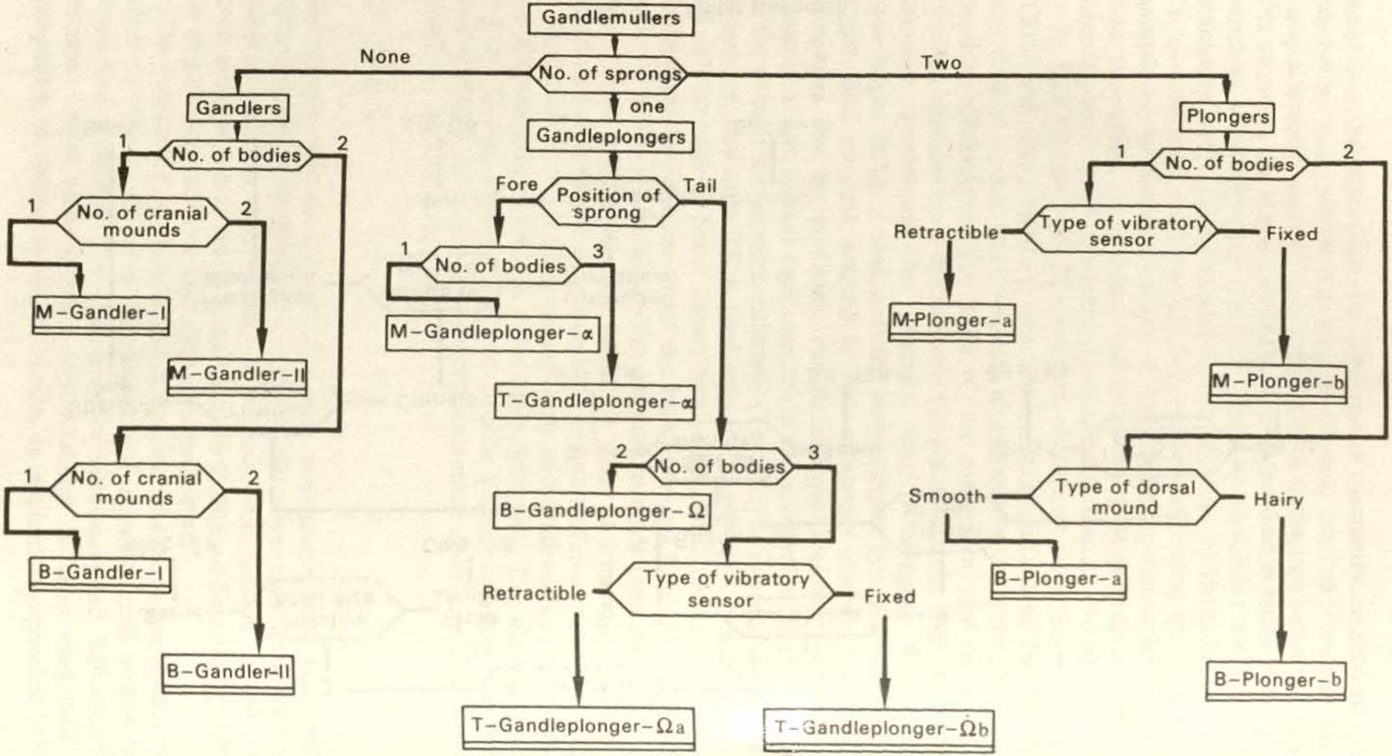


Fig. 5 Gandlemuller taxonomy

subject competence's quite accurately from this information. The authors can both do so, though we have only the most general concept of the cues we employ for this purpose. In fact, the following subjects were non-randomly selected and their competence types accurately predicted.

(*a*) *Serialists.* 2 electrical engineering students engaged on programming 1 mathematics graduate, 1 economics graduate and 1 physics graduate.
(*b*) *Holists.* 1 philosophy graduate currently working in social psychology, 1 physics graduate studying operational research, 2 sociology undergraduates.

2.4 *Session (I) Free learning and description of the task (I) universe*

Information about the (free learning) Task (I) universe is written on cards that belong to five classes A, B, C, D, E. On the front of each card is a statement which specifies the type of data to be found on its reverse (or face) side. The cards are arranged face downwards in a five-column array in which each column contains the members of one of the classes A, B, C, D, E. Every column is labelled to indicate the *class* of data written on any of the member cards, namely:

Class A 10 cards showing pictures of a typical member of a Clobbit subspecies;
Class B 13 cards giving contextual information about the appearance, habitat, or behaviour of these animals;
Class C 5 cards stating the structure of the taxonomy. Each card names the members of a subspecies and states the numbers of physical characteristics tested in order to distinguish this subspecies;
Class D 7 cards naming physical characteristics of the clobbits and the values assumed by each property on each of the subspecies which it is used to distinguish;
Class E 7 cards indicating the subspecies names or codes and why they were used.

Prior to free learning the student is oriented by reading through an introduction which states the goal and *all* of the strictly relevant information. After this he is allowed 25 minutes in which to ask questions. To ask a question the student selects one card at once, and turns it over to scrutinize the message on its face. During this process he knows (as noted in section 1.2) that the cards convey redundant information, i.e. that all the requisite data could be obtained from A and C; from A and D; from C, D, and E; or from B and certain subsets of the remaining cards in the array.

For each question (or, in the case of a coherent search, each sequence of questions) the student is required to give a reason for asking the question concerned and his reasons are tape-recorded. The time occupied in giving reasons is not deducted from the time allowed for questioning. As noted in section 1.2 only four sorts of reason were actually given (to recapitulate): (i) search a card class or co-ordinate of the message space; (ii) search for a specific datum; (iii) test a single predicate hypothesis; (iv) test a multiple predicate hypothesis.

Of these, the occurrence frequency of type (iv) reasons is used to discriminate serialist and holist competence, for which purpose it is unnecessary to enquire *what* hypothesis is being tested. In general (for example, for the questioning languages of section 1.5 and 1.6) it would be necessary to delineate the hypotheses; but they are readily specified within the framework afforded by the array and the four categories of reasons for questioning.

It is predictable, from the theory, that holist mentation will be discriminated from serialist mentation by the occurrence, in free learning or in teachback of tests for hypotheses based on complex predicates rather than simple ones. Due to the unitary nature of the tests in the classification tree a simple predicate hypothesis is probably a single predicate hypothesis and, given this identification, the discrimination is clearcut. Table 1, is the 'analysis of reasons' on which the initial assignment of competence type was based.

A more comprehensive picture of free learning is given by the four typical learning strategy graphs of Appendix 1. Each of the main nodes

Table 1 *Molecular data from free learning session (1): Competence Type assignment based on type (of predicate) in hypothesis*

		Information seeking transitions	*'Small' predicate in hypotheses*	*'Large' predicate in hypotheses*
Serialist	1	19	3	0
	3	10	7	0
	4	37	0	0
	6	20	0	0
	8	9	9	0
	10	26	6	0
	12	18	1	0
	16	18	5	0
Holist	2	10	8	8
	5	21	2	13
	7	22	8	5
	9	8	2	14
	11	8	2	10
	13	7	5	9
	14	11	0	8
	15	20	0	5

represents a note made by the student. The transitions between nodes represent the order in which the student elected to ask questions and the transitions are coded to depict his reasons for doing so.

APPENDIX 1

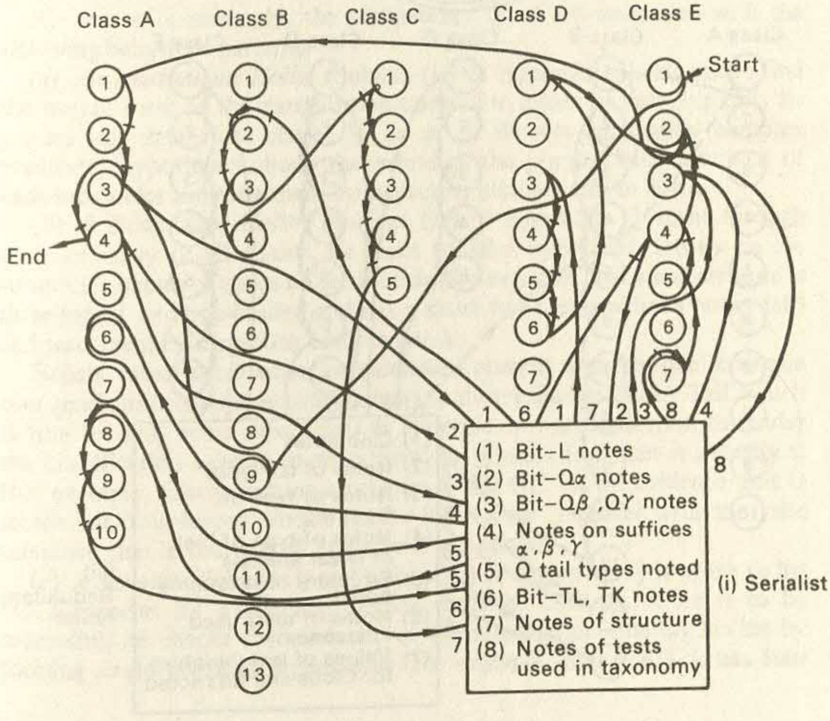
Typical Learning Strategy Graphs

(i) and (ii) serialist students
(iii) redundant holist
(iv) irredundant holist

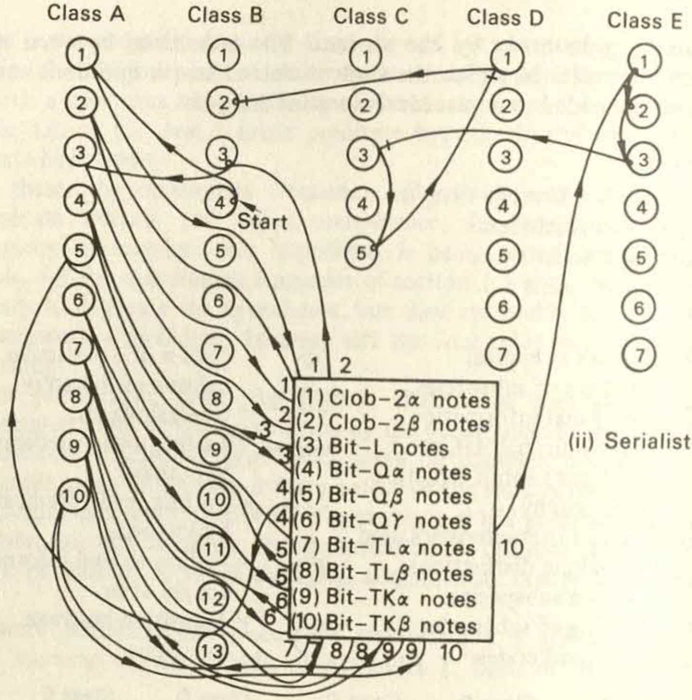
KEY

Class A	pictures of typical members of subspecies.
Class B	contextual information (behaviour, habitat, etc.)
Class C	statements about structure of taxonomy.
Class D	physical characteristics used for making distinctions between subspecies.
Class E	meaning of subspecies names and codes.

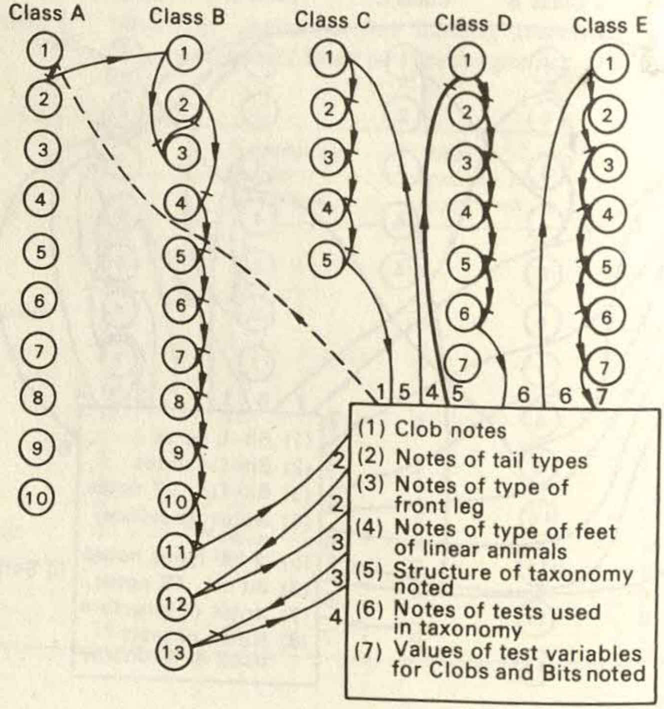
ⓝ Card *n* of a particular column of the array.
—— = General search.
━━ = Testing many predicate hypothesis.
- - - - = Testing single predicate hypothesis.
↘○ = Failure to find information sought after.
↘○ = Transition between questions.



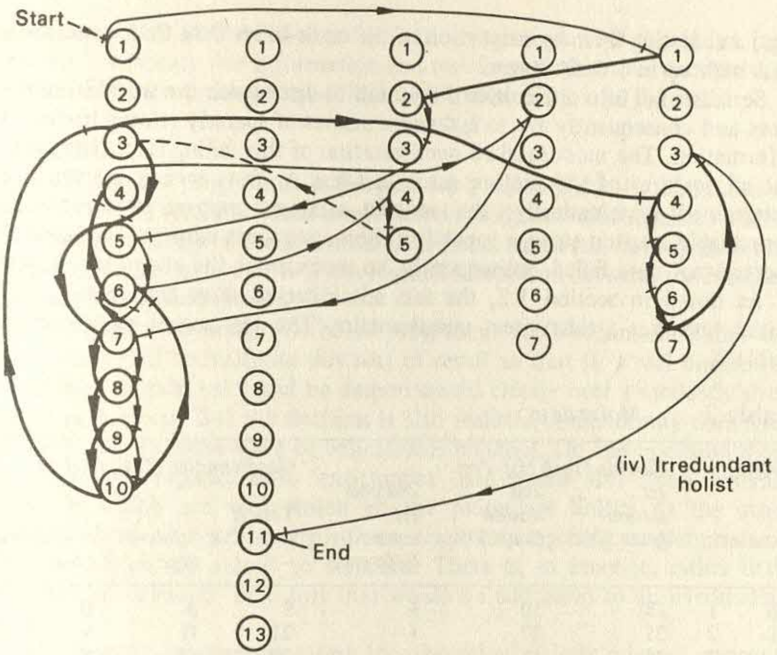
(i) Serialist



(ii) Serialist


Class A
Class B
Class C
Class D
Class E
(1) Clob notes
(2) Notes of tail types
(3) Notes of type of front leg
(4) Notes of type of feet of linear animals
(5) Structure of taxonomy noted
(6) Notes of tests used in taxonomy
(7) Values of test variables for Clobs and Bits noted

(iii) Redundant holist



By way of a summary, the competence types are associated with the following behaviour patterns:

(*a*) *An Irredundant Holist* [Subject (iv) of Appendix 1] sets out to find the overall form of the classification tree and to locate its tests; typically he garners this data from classes B, D, or E. He sets up (usually complex predicate) hypotheses about the values of the physical characteristics of each subspecies and tests them by inspecting picture cards in Set A.

(*b*) *A Redundant Holist* [Subject (iii) of Appendix 1] scans through cards of many or all classes. He bases complex hypotheses mostly on the subspecies names or codes of Set E cards (for example, that some creature is three-legged and one-headed and has a short front leg and has a bushy tail) and tests them by inspecting cards in Set A.

Holists, either *irredundant* or *redundant* commit mistakes due to simple over-generalization (for example, that (β) always implies 'Bushy Tail' which is true for only *some* subspecies) or systemic over-generalization to render the classification *scheme* more rational or symmetrical than it actually is (for example, falsely naming a subspecies 'Bit QL' on the evidence that Q stands for 'four-legged' and L stands for 'Linear' together with the valid inference that a four-legged linear creature exists).

(*c*) *A Serialist* [Subject (i) and subject (ii) of Appendix 1] is prone to list the subspecies by examining picture cards in Class A. If he is to be successful, he checks the relevance of the information entering his list by forming single predicate hypotheses (for example, that a BIT L has four legs) and testing them by inspection of the cards in set D or C. A structure is thus built up in orderly stages.

Serialists fall into difficulties if they fail to distinguish the wood from the trees and consequently try to assimilate masses of sparsely related irrelevant information. The most marked manifestation of this failing is an attempt to list all features of the picture cards in Class A. But, even if relevance is determined systematically, the serialist strategy appears to impose an appreciable load on storage capability. Some students who accumulated *all* the relevant data failed, subsequently, to *reconstruct* the entire taxonomy.

As noted in section 1.2, the test administered after free learning the initial task was a thirty-item questionnaire. The test scores are shown in Table 2.

Table 2 *Molar data*

Subject Number	*Clobbit [Task (I)] Test 1st session (max 30)*	*2nd session (max 30)*	*2nd and 1st score difference*	*Gandlemuller (Task [II)] test Task II test*	*Student type*	*Programme type*
TB 1	13	30	+	9	S	H
TB 2	25	27	+	21	H	S
3	25	17	–	29	S	S
TB 4	10	28	+	7	S	H
5	27	23	–	30	H	H
6	15	10	–	9	S	H
TB 7	27	30	+	30	H	H
8	27	18	–	28	S	S
TB 9	28	30	+	30	H	H
10	17	13	–	30	S	S
11	23	19	–	19	H	S
TB 12	18	30	+	9	S	H
13	21	25	+	30	H	H
14	26	17	–	16	H	S
TB 15	21	28	+	20	H	S
16	22	17	–	29	S	S

2.5 *Session* (II) *Procedures*

Achievement at Task (I) was again tested at the beginning of Session (II). The test scores are shown in Table 2.

Task (II) is to learn the taxonomy of Figure 5 on the universe discussed in section 2.2 This universe is also redundantly described by a set of cards. These were used to generate the serialist and holist teaching programmes employed for instructing Task (II). Both of the programmes convey exactly the same *strictly relevant* information, knowledge of which is *subsequently tested.* Both programmes have a linear presentational format (the serialist programme has eleven information and test frames and the holist programme has twenty five information and test frames). But the concepts they instruct are different and very differently structured. [. . .]

APPENDIX 2

. . . The holist programme is designed for a *redundant* holist and scrutiny of the Task (II) teachback data shows that it wholly defeats a serialist student. A programme designed for an *irredundant* holist might not have done so and thus might have given rise to a less marked distinction between the matched and the mismatched groups.

The decision to design the holist programme for a redundant holist was partly intended to facilitate this sort of result so that (if a real distinction was found to exist) it could be demonstrated clearly over a modestly sized group of subjects. But the decision is also realistic. A dichotomy does exist between two extreme sorts of educational material. On the one hand there are materials replete with 'enrichment' and based also upon 'overall' concepts which are well suited to the redundant holist. At the other extreme there are teaching algorithms, single step teaching programmes and the like which are suited to serialists. There is, in practice, rather little material 'in between' or a sort that would be addressed to an irredundant holist.

The serialist programme refers to a kernel of strictly relevant concepts which are presented in an orderly and logical fashion. Inspection of the Task (II) teachback shows that it defeats a redundant holist and offers a difficult (and incidentally, unpalatable) learning experience to an irredundant holist, though students with this competence do eventually come to grips with it.

Both programmes were administered in the same way with the following instructions and criteria for advancement. Recall that both programmes are in a linear format, and thus consist in a series of information and test frames.

1. At the *n*th frame, $n = 1$, the subject is asked to read slowly through the information frame (actually, a card of the sort used for questioning and answering during the free learning task).
2. When the subject has finished, he is presented with a test frame (a card containing questions and blank space to fill in).
3. If the subject responds correctly, he is presented with the $n + 1$th information frame (unless the *n*th frame is the last, when step 6, below, is executed).
4. If the response is mistaken the *n*th information frame is presented and read once again.
5. The cycle is repeated so that a subject does not advance to the $n + 1$th frame until his response to the *n*th is correct in all particulars.
6. On completion of the programme, the student either has or has not completed an error free run. If he has done so, the procedure terminates.

7. If any error has been committed the entire programme is repeated until such moment as the criterion stated in stage 6 is satisfied.
8. Since the procedure is cyclic, all students are eventually forced to complete one run of the programme to which they are assigned without any overt mistakes.

Let us emphasize that satisfaction of this criterion does not imply that a student has learned the taxonomy even though he gives the appearance of mastery and even though criteria of this type are often believed to indicate mastery is attained.

Next, a student's knowledge is tested by means of a thirty-item questionnaire. The test scores obtained from holists and serialists who received instruction via a matched mode (holist/holist programme; serialist/serialist programme) or via a mismatched mode (holist/serialist programme; serialist/holist programme) are shown in Table 2.

The final event in Session (II) was the elicitation of 'teachback' (section 1.2) the results of which were recorded and content analysed.

(i) To check the original competence classification, based on the free learning of Task (I) during Session (I).
(ii) To elucidate the processes involved in matched and mismatched learning.

Some typical 'Teachback' protocols are set out in Appendix 3. The differences between students and cognitive dilemmas they encountered are apparent from an inspection of this data. To quantify them, two sorts of analysis were conducted, the results of which are shown in Table 3 and Table 4.

APPENDIX 3

Typical Teachback Protocols for Task II

KEY

n→ ←*n*	Information derived from frame *n*
━━━━━━	Inventions
──────	Irrelevant information
▬▬▬▬▬▬	Redundant information
- - - - - -	Statement referring to information to be presented or previously mentioned
∿∿∿∿∿∿	Information inferred from the programme by the subject.
══════	A false statement.
+	Cross reference of a false statement with a correction.
[ ]	Repetitions.
( )	Other comments, interjections, etc.

(i) *Holist with holist programme*

(1) I'm going to tell you about a $^{1\to}$funny Martian animal which has been recently discovered and classified by scientists conducting surveys$^{\leftarrow 1}$ They are funny slug-like things with $^{1\to}$various protruberances, some of which differ amongst the different types. Zoologists use these differences to draw up a classification$^{\leftarrow 1}$ $^{2\to}$ There are other ways of telling them apart$^{\leftarrow 2}$ $^{7\to}$The main alternative is a Russian scheme which uses masticator differences where the standard scheme talks about sprongs$^{\leftarrow 7}$ $^{8\to}$which are a kind of horny spike, used for defence against predators, notably the Owzard or night vulture which is also Martian$^{\leftarrow 8}$

(2) $^{1\to}$These animals are called Gandlemullers,$^{\leftarrow 1}$ $^{3\to}$because they churn about in the swamps near the equator and Gandle is Martian for swampmud, hence swampmudmiller–muller is German for Miller$^{\leftarrow 3}$. [These things churn through the mud], $^{9\to}$eating it by some curious process which means they eat and excrete at the same time ^{9}–like a flow through bakery.

(3) [Anyway, they sit about eating this stuff] and all $^{18\to}$the time they wave tendrils in the air called vibratory sensors for detecting Owzard's wing beats$^{\leftarrow 18}$. When attacked, they do various things depending on their defence equipment. For a start $^{19\to}$some of them retract their sensors. Some don't because they are fixed$^{\leftarrow 19}$ $^{10\to}$Some of them lie all flat and squashed because they have no spikes for stabbing.$^{\leftarrow 10}$ Others have these sprong things and do various gymnastics. If $^{11\to}$they have a front and back spike they jackknife.$^{\leftarrow 11}$ If $^{13\to}$they have only a front one, they rear up on a thing at the back called a ventral skirt$^{\leftarrow 13}$. If $^{14\to}$their spike is at the back, they do it the other way on a thing called a foreplank$^{\leftarrow 14}$ which, by the way, $^{10\to}$ is collapsed by the ones that have no sprongs so they can get flat$^{\leftarrow 10}$

(4) Anyway, if we come to the classification proper you'll see how they differ in the various ways.

(5) (Now) $^{1\to}$There are 13 different subspecies$^{\leftarrow 1}$ $^{6\to}$divided into 3 main types. There are $^{15}\leftarrow$4 Gandlers,$^{\to 15}$ $^{18\to}$4 Plongers$^{\leftarrow 18}$ and $^{23\to}$5 Gandleplongers$^{\leftarrow 6,\ 23}$.

(6) You $^{5\leftarrow}$get these with various numbers of bodies–one, two or three and a subspecies has a prefix M(mono), B(bi), T(tri) to say which is which$^{\to 5}$.

(7) There $^{15\to}$are 2 M-Gandlers, 2B-Gandlers.$^{\leftarrow 15}$. Similarly there $^{18\to}$are 2 M-Plongers and 2 B-Plongers$^{\leftarrow 18}$

(8) Now $^{5\rightarrow}$in Gandleplongers you get the T-types$^{\leftarrow 5}$ Oh, $^{12\rightarrow}$there are 2 sorts of Gandleplonger called $^{13\rightarrow}\alpha$ or Ω. $^{14\rightarrow}$depending on their sprong. You see they have only one sprong$^{\leftarrow 12}$. α is the one with it at the front$^{\leftarrow 13}$ and Ω is at the rear$^{\leftarrow 14}$ They're the ones that pivot at the front or back depending, as I said before. Anyway, $^{23\rightarrow}$there are 2 α's an M and a T and there are 3 Ω's, one B and two T's$^{\leftarrow 23}$. I'll say how you tell them apart in a minute, but first I must say about sprongs.

(9) $^{10\rightarrow}$Gandlers are the ones with no sprongs$^{\leftarrow 10}$ $^{11\rightarrow}$Plongers are the ones with 2 sprongs$^{\leftarrow 11}$ and $^{12\rightarrow}$[Gandleplongers have one at the front or back$^{\leftarrow 12}$, as I just said$^{13\rightarrow}\alpha$ and $^{14\rightarrow}\Omega^{\leftarrow 13,14}$].

(10) Now I'll do Gandlers. You $^{15\rightarrow}$have to tell the two M's from the B's. This is by cranial mound which is a bump on the head. The MI and BI have a single bump. The MII and BII have a double bump which $^{16\rightarrow}$shows they are very clever. So I means single, II means double$^{\leftarrow 15,16}$

(11) Now Plongers. $^{18\rightarrow}$There's an Ma and Mb and a Ba and Bb$^{\leftarrow 18}$ Here we use two different tests. One of them we use again in a minute.

(12) Now $^{18\rightarrow}$the M-Plongers are distinguished by the vibratory sensor and you find that the M-Plonger-a has a fixed sensor^{+} (no sorry, retractible^{+}), whereas the Mb is fixed$^{\leftarrow 18}$. Turn now to the B's$^{20\rightarrow}$. Here we find that one of them has a hairy dorsal mound. Hairiness is a good thing because it's an insulation against heat loss in digestion.

(13) The one with the hairy mound is B-Plonger-b$^{\leftarrow 20}$. Anyway that's the Plongers, [the prickers with 2 sprongs.] The last ones are the Gandleplongers.

(14) [As I said there are 2 sorts of α and Ω's. [$^{23\rightarrow}$The α's are Mα and T$\alpha^{\leftarrow 23}$.] That's all we need to know number of bodies and sprong type. Now $^{23\rightarrow}$we have one BΩ on its own. And two TΩ's$^{\leftarrow 23}$. There $^{24\rightarrow}$are Ωa and Ωb there we find a difference we used before. It's vibratory sensor and again its the a which is retractible. The b is fixed$^{\leftarrow 24}$ (That should do).

(ii) *Serialist with serialist programme*

(1) $^{1\rightarrow}$Zoologists have classified the Gandlemuller on the basis of physical characteristics$^{\leftarrow 2}$

(2) $^{2\rightarrow}$The three main types are Gandlers, Plongers and Gandleplongers$^{\leftarrow 2}$. $^{3\rightarrow}$Gandlers have no sprongs. Plongers have two sprongs. Gandleplongers have one sprong$^{\leftarrow 3}$

(3) $^{4\rightarrow}$There are four subdivisions of Gandler; MI, MII, BI and BII$^{\leftarrow 4}$ $^{5\rightarrow}$The M's have one body, the B's have two bodies. The MI and BI have a single cranial mound. The MII and BII have a double cranial mound.$^{\leftarrow 2}$

(4) $^{6\rightarrow}$There is an M-Plonger-a and b and B-Plonger-a and B$^{\leftarrow 6}$ ^{7} M's with one body and are distinguished by the type of vibratory sensor, which for M-Plonger-a is retractible. The b is fixed.

(5) The two B-Plongers have two bodies and are distinguished by type of dorsal mound. The B-Plonger-a has a smooth dorsal mound, the 'b' has a hairy one$^{\leftarrow 7}$

(6) (Let's see,) $^{8\rightarrow}$there are 2 Gandleplonger-α's and 3 Ω's$^{\leftarrow 8}$

(7) ^{9} The α's are the Mα and the Tα with one and three bodies respectively$^{\leftarrow 0}$

(8) ^{10} The Ω's have one B and two T types, TΩa and TΩb $^{\leftarrow 10}$.

(9) (Oh) $^{8\rightarrow}$α's have a foresprong, Ω's have a tailsprong$^{\leftarrow 1}$ and $^{11\leftarrow}$the T-Gandleplonger Ωa has a retractible vibratory sensor whereas the b is fixed$^{\leftarrow 11}$

(iii) *Holist with serialist programme*

(1) Now, $^{2\rightarrow}$there are 3 main types of $^{1\rightarrow}$Gandlemuller$^{\leftarrow 1}$. $^{4\rightarrow}$There are four called Gandlers$^{4\leftarrow}$, $^{6\rightarrow}$four called Plonger$^{\leftarrow 6}$ and$^{8\rightarrow}$ 5 called Gandleplonger$^{8\leftarrow}$. That's 13 altogether.

(2) The group of 5 come in two kinds α or Ω depending on a physical characteristic$^{\leftarrow 8}$ (Oh yes) $^{1\rightarrow}$Zoologists have made a classification based on physical$^{\leftarrow 1}$ characteristics.

(3) $^{2\rightarrow}$Some of them have spiky things, some don't$^{\leftarrow 2}$.

(4) Some have 1 body, some have 2 or 3.

(5) You also have various hairy or smooth, single or double mounds.

(6) There's something else. Retractible vibratory sensors, or fixed. That's used twice.

PAUSE

(7) (Let's see). $^{4\rightarrow}$There are 4 Gandlers, M and B$^{\leftarrow 4}$ $^{6\rightarrow}$4 Plongers M and B$^{\leftarrow 6}$ and $^{8\rightarrow}$5 Gandleplongers$^{\leftarrow 8}$ $^{9\rightarrow}$M & T of one sort$^{\leftarrow 9}$ and $^{10\rightarrow}$M and B of other$^{\leftarrow 10}$ (No, B and T^{+}) You $^{4\rightarrow}$tell the Gandlers apart by a suffix I or II$^{\leftarrow 4}$ telling you whether $^{5\rightarrow}$they have this mound single or double$^{\leftarrow 5}$

(8) You tell ^{6} $^{7\rightarrow}$Plongers apart by sensors, a fixed b retractible ^{6}, $^{7\rightarrow}$(It might be the wrong way there) (I think) $^{7\rightarrow}$you also get the mound in here. Hairy is b$^{\leftarrow 7}$

(9) The $^{8\rightarrow}$Gandleplongers have a thing which is fore or aft. You put α or Ω after the name$^{\leftarrow 8}$

(I've missed something, but I can't really remember all the bits. I got quite a lot though.)

(iv) *Serialist with holist programme*

(1) In $^{1\rightarrow}$surveys of Mars, Zoologists have classified Gandlemullers into 13 subspecies depending on physical characteristics$^{\leftarrow 1}$.

(2) $^{2\rightarrow}$There are also differences in behavioural habits$^{\leftarrow 2}$.

(3) $^{3\rightarrow}$Gandle is mud. Muller is miller$^{\leftarrow 3}$.

(4) They $^{4\rightarrow}$have various physical things that are used . . . There are sprongs, foreplank and a ventral thing. Masticators, an aperture and a dorsal mound$^{\leftarrow 4}$.

PAUSE

(5) $^{6\rightarrow}$There are 3 main subspecies: M $^{5\rightarrow}$means mono, B means Bi and T means tri, for number of bodies$^{\leftarrow 5}$. The names are Gandler, Plonger and Gandleplonger$^{\leftarrow 6}$ $^{3\rightarrow}$They live in caves$^{\leftarrow 3}$.

(6) $^{10\rightarrow}$Gandlers have no sprongs$^{\leftarrow 10}$ $^{11\rightarrow}$Plongers have 2$^{\leftarrow 11}$ and $^{12\rightarrow}$Gendleplongers have 3$^{\leftarrow 12}$.

(7) [$^{8\rightarrow}$They have masticators] for eating mud$^{\leftarrow 8}$ $^{9\rightarrow}$which takes 2½ minutes and is stored$^{\leftarrow 9}$.

(8) $^{10\rightarrow}$When Owzards, the Martian night vulture, attacks Gandlemullers, he lies flat excreting all mud$^{\leftarrow 10}$. (No, that's just Gandlers)$^{+}$ because $^{11\rightarrow}$Plongers prick or stab and jacknife$^{\leftarrow 11}$.

(9) Some have $^{12\rightarrow}$spoon-shaped masticators which are hairy or smooth$^{\leftarrow 12}$.

(10) You $^{15\rightarrow}$get M-Gandlers and B-Gandlers with suffix I and II meaning single or double cranial mound$^{\leftarrow 15}$.

(11) You get $^{18\rightarrow}$q or Ω meaning retractible or fixed sensor$^{\leftarrow 18}$. (I can't remember any more with much clarity). I know some of them $^{25\rightarrow}$swim in various ways$^{\leftarrow 25}$. And its a $^{20\rightarrow}$B-Plonger-b that has a hairy something$^{\leftarrow 20}$. and $^{24\rightarrow}$there are TΩa's and b's depending on the number of spikes$^{\leftarrow 24}$.

The first analysis (Table 3), is a content analysis of the number of occurrences of:

(*a*) falsehoods;
(*b*) corrected falsehoods;

Table 3 *Content analysis of teachback protocols (Task II)*

Subject No.		*Falsehoods/corrected falsehoods*	*Inventions*	*Statements of information to come or delivered*	*Statements deduced*	*Irrelevant*	*Redundant*	*No. of statements—total (excludes interjections, repetition or corrections)*			
Serialist with serialist programme	9	0	0	0	0	0	0	29	Serialist	Serialist material	Matched students
	8	1/1	0	0	0	0	0	28			
	16	1/1	1	1	1	0	0	41			
	3	1/0	0	3	1	0	0	47			
Holist with serialist programme	14	1/1	0	0	7	0	0	37	Holist		Mismatched students
	11	2/1	0	8	4	1	0	33			
	2	0	1	2	7	1	0	26			
	15	3/1	0	1	4	0	0	34			
Serialist with holist programme	1	2/0	0	1	0	11	2	39	Serialist	Holist material	
	12	3/1	0	0	0	7	0	28			
	6	4/1	0	0	0	14	3	42			
	1	6/0	1	0	1	20	4	48			
Holist with holist programme	5	0/0	1	5	3	9	0	66	Holist		Matched students
	9	1/1	1	12	5	5	0	70			
	7	1/1	2	13	1	23	3	110			
	13	2/1	6	17	10	23	24	116			

(*c*) inventions (items imported from outside the described universe);
(*d*) anticipatory statements about what is to come later;
(*e*) deductive statements;
(*f*) irrelevant statements (i.e. those that are irrelevant to the required body of knowledge;
(*g*) redundant statements (i.e. those that over-specify this body of knowledge);
(*h*) the total number of statements analysed for a given subject.

The second analysis (Table 4) is a teachback order analysis intended to show the extent to which a student's self-constructed teachback of the concepts he has learned departs from the order in which these concepts were presented by the teaching programme to which he was assigned. The tabulated numbers are frame numbers in the respective teaching programme, (holist or serialist). The order of the tabulated numbers is the sequence in which the pertinent concepts were evoked during teachback.

Table 4 *Teachback order in relation to programme frame no.*

	Subject No.	
Serialist with serialist programme	9	1→2→3→4→5→6→7→8→9→10→8→11
	8	1→2→3→4→5→6→7→6→7→8→9→10→11
	16	1→2→3→4→5→6→7→8→9→10→11
	3	1→2→3→4→5→6→7→8→9→10→11
Holist with serialist programme	14	2→1→4→6→8→1→2→4→6→8→9→10→4→5→$\substack{6\\7}$→8
	11	1→2→1→9→2→3→4→5→7→11→8→$\substack{9\\10}$→11
	2	1→2→4→6→8→2→8→2→3→8→1→7→5→7→4→6
	15	1→2→1→2→3→4→6→8→4→5→6→7→9→10→9→11
Serialist with holist programme	1	3→4→6→10→11→12→7→8→9→15→9→21→11→13→14→23→24→25
	12	1→2→3→8→4→5→6→10→11→12→13→8→15→16→18→20→24
	6	1→2→3→4→6→5→3→10→11→12→8→9→10→11→12→$\substack{15\\20}$→$\substack{18\\24}$→25
	4	1→3→5→4→$\substack{6\\12}$→11→15→16→20→9→16→18→11→13→14
Holist with holist programme	5	1→3→1→6→5→4→15→16→18→20→13→14→7→15→18→20→18→15→20→23→24→13→10→11→12→13→14→24
	9	1→2→1→5→7→8→10→11→12→$\substack{13\\14}$→5→15→18→22→18→20→23→22→24
	7	1→2→7→8→1→3→9→18→19→10→11→13→14→10→1→6→15→18→23→5→15→18→5→12→13→14→23→10→11→12→13→14→15→16→18→20→23→24
	13	1→6→5→7→8→10→3→7→11→12→6→5→15→18→10→15→4→16→15→10→16→15→6→18→20→22→19→23→12→13→14→23→5→24

2.6 *Discussion of the numerical results*

Scrutiny of Table 1 [Task (I) free learning data] indicates a clearcut distinction between serialists and holists in terms of 'complex' or 'large predicate' hypotheses. The distinction can be heightened by taking the nature of the hypotheses into account and, in this case, it is also evident that the *type* of single predicate hypotheses tested by the serialist differs from the *type* (of single or complex hypotheses) tested by the holist. There is also a positive correlation between success and the total number of hypotheses tested.

The same competence assignment can be made in terms of the Task (4) teachback data (Tables 3 and 4). In Table 3 (content analysis) holists are distinguished from serialists in terms of the number of inferential statements they produce where an inferential statement is *either* an *anticipatory* statement *or* a *deductive* statement. In Table 4 (teachback order analysis) it is possible to distinguish the serialist from the holist by a tendency, on the part of a serialist, to preserve the order of the programme presentation format which is absent in the holist. Presented with a holist programme the serialist is unable to preserve the complete order but he does manage to preserve sequentially arranged fragments. So far as we know there is no generally accepted statistical test for similarity matching discrete patterns of variable length. Hence, no analysis is offered.

The gross data (test scores for holists and serialists, matched and mismatched) are shown in Table 2. First, it is quite evident, on examining the Task (II) test scores, that matched students learn more efficiently than mismatched students. Since the rankings do not overlap a statistical test is gratuitious but, as a formality, the difference between matched students $N = 8$ and mismatched students $N = 8$ is significant (in favour of the matched group) at the 0.1 per cent level of significance ($0.001 > P$) using the Mann-Whitney U-test. The same comments apply to an intergroup comparison of the ratios:

$$\frac{\text{Task (II) Test Score (Session II)}}{\text{Task (I) Test Score (Session I)}} \times 100\%$$

Comparing the ($N = 4$) mismatched holists with the ($N = 4$) mismatched serialists there is a difference in favour of the holists, at the 1 per cent level ($0.01 > P$) which we believe to reflect the fact that an irredundant holist can (with difficulty) learn from a serialist programme. In fact, those mismatched subjects classified as irredundant holists learned the task after 4-7 repetitions of the serialist teaching programme, compared with the 1-3 repetitions required by (matched) serialists. As desired (condition 4 of section 2.1) there is no significant difference between the scores achieved by matched serialists and matched holists.

The influence of teachback on learning Task (I) is reflected in the test scores achieved for Task (I) at Session (II) [after the teachback experience of Session (I)]. Comparing the scores of the ($N = 8$) students receiving a real teachback experience (marked TB) with those of the ($N = 8$) students receiving ineffective teachback there is, once again, no overlap in the ranking and the bias is strongly in favour of teachback.

The difference in question is significant at the 0.1 per cent level ($0.001 > P$) using the Mann-Whitney U-test.

2.7 *Peripheral results*

From the theory, we predict a positive correlation between the number of *hypotheses* posed by a student and his efficiency. The appropriate figures to examine are 'number of hypotheses per unit time', this point is supported by the data of Table 1.

Turning to Table 3, it is possible to distinguish between the effect on teachback of programme type (serialist/holist) and student competence (serialist/holist). Relative to a serialist programme, the holist programme injects either *irrelevant* or *redundant* statements into the teachback. Relative to a serialist student, a holist student is prone to make inferential statements. The holist comes into his own in the matched condition when he can *use* the irrelevant or redundant material provided by a holist programme in order to make *inferences*.

The teachback data of Table 4 deserves further scrutiny, since it might lead to a very objective method for assigning competence types. Given a serialist programme, the serialists show a remarkable tendency to preserve the concept (and frame) ordering. In contrast, the holists disrupt it quite effectively. It should be noted that holists have about the same disrespect for the ordering inherent in a holist programme. The weak point is the difficulty (without an appropriate matching test) of demonstrating that serialists *do* try to preserve the ordering of a holist programme, even though, almost by definition, their endeavours will be thwarted.

2.8 *Auxiliary study*

As noted in section 1.3 a subgroup of the competence assigned subjects learned the 'operon' task, described in section 2.1 under matched and mismatched conditions (i.e. as before, using a teaching programme that did or did not suit their competence). No molecular data was collected but the molar results (test scores for a test administered after learning) are shown in Table 5.

Table 5 *Test scores on auxiliary task ('System Learning')*

Subject	*Test score maximum = 20*		
10	20	Serialist subjects	Serialist programme
3	17		
6	18		
8	6		Holist programme
16	8		
4	10		
2	11	Holist subjects	Serialist programme
15	9		
14	13		
13	20		Holist programme
7	19		
11	19		

Judging from this data, a competence assignment based on taxonomy learning may be extrapolated into the area of 'system learning'. If students are matched to their instruction, according to the original assignment, they learn effectively; if they are mismatched they fail to do so. There is enough circumstantial evidence for differences in the mechanism of 'system learning' under matched and mismatched conditions to justify the collection and analysis of the molecular data required to support this point of view.

References

JACOB, F. and MONOD, J. (1961) 'On the regulation of gene activity', in *Cold Spring Harb. Symp. Quant. Biol., 26, Cellular Regulatory Mechanisms.*

PASK, G. (1965) 'Comments on the organization of men, machines and concepts'. *In Education for Information Science.* Ed. L. HEILPRIN et al. p. 133. New York: Spartan Press; London: Macmillan.

PASK, G. (1968) 'A cybernetic model for some types of learning and mentation.' Bionics Symp., Dayton, Ohio, 1966. In *Cybernetic Problems in Bionics.* Ed. H. C. OESTREICHER and D. R. MOORE, p. 531. New York, London: Gordon & Breach.

PASK, G. (1969a) 'Strategy, competence and conversation as determinants of learning', *Programmed Learning,* Oct. 1969, p. 250.

PASK, G. (1969b) 'Adaptive machines.' *Proc. NATO Symp. Major Trends Programmed Learning Res.,* p. 251. Paris: Dunod.

PASK, G. (1969c) 'Interaction between a teaching machine and the student's attention directing systems.' *Proc. 16th Int. Congr. Appl. Psychol.,* p. 269. Amsterdam: Swets & Zeitlinger.

PASK, G. (1970a) 'Computer assisted learning and teaching.' Paper presented at NCET Conference on Computer Based Learning, Leeds.

PASK, G. (1970b) 'Fundamental aspects of educational technology (illustrated by the principles of conversational systems).' IFIP World Conference on Computer Education. Conference edition of papers in three parts. Ed. B. SCHEEPMAKER, Pt. 1. Invited papers.

PASK, G. (1971) 'Teaching machines.' In *Modern Trends in Education.* Ed. B. ROSE. London: Macmillan.

PASK, G. and LEWIS, B. N. (1968) 'The use of a null-point method to study the acquisition of simple and complex transformation skills', *Br. J. math. stat. Psychol., 21,* 61.

PASK, G. and SCOTT, B. C. E. (1969) Final scientific report. USAF Contract No. F61-052-67-C-0010.

PASK, G. and SCOTT, B. C. E. (1971) 'Learning and teaching strategies in a skill transformation,' *Br. J. math. stat. Psychol., 24,* 205.

UPTON, A. and SAMPSON, L. W. (1963) *Creative Analysis.* New York: Dutton.

21 Some implications of research on cognitive style for problems of education

H. A. Witkin

For some time our laboratory has been engaged in research on the relation between an individual's personality and his cognitive functioning–that is, his perceptual and intellectual activities. A particular focus of interest in this work has been the characteristic, self-consistent ways of functioning a person shows in the cognitive sphere. We call these his 'cognitive styles'. Research on cognitive styles has progressed to a point where it may be profitable to inquire into its implications for problems of education. The concepts and findings that have emerged from his work may find useful application to problems of selection, placement and methods of teaching. In particular, this work is likely to contribute to a theoretical base for approaching these problems and to open up significant lines of research for attacking them.

This paper points out some of the places where connections may exist between the work on cognitive style and issues of education. Educators, familiar with school problems, are, of course, in the best position to judge the merit of these ideas and to decide about which connections it may be fruitful to pursue. In this attempt the particular cognitive style with which we have ourselves been concerned will be considered as illustrative. To explain the nature of this cognitive style, a brief account will first be given of the work, as it unfolded, in which this cognitive style was identified and its relation to personality functioning and early life experiences explored (Witkin *et al.*, 1954; Witkin *et al.*, 1962).

Individual differences in cognitive style

Very early in this work we found that people differ from each other in their characteristic way of perceiving both the world and themselves. We also found that a given way of perceiving is related to the personality adjustment the individual has worked out in the course of growing up. Perhaps the best way to describe the individual differences in mode of perceiving is to tell about the actual test situations we use to identify them.

Source: Chapter 9 in GOTTESGEN, M. B. and GOTTESGEN, G. B. (eds.) (1969) *Professional School Psychology*, Vol. 3. New York: Grune and Stratton, 198–227.

Individual differences in perception: field-dependence-independence

Among these situations are some concerned with the very commonplace perceptual function of orientation toward the upright in space. As is well known at every moment, except perhaps sometimes in an airplane or an amusement park, we have a quick, accurate sense of whether we are straight or tilted and whether objects around us are straight or tilted. Behind this smooth-functioning process lies a rather complicated mechanism. Obviously, one way we can tell that we are straight is that we feel straight: and we can judge the position of other objects in relation to the felt position of our bodies. In addition, the world in which we live is filled with prominent verticals and horizontals, perceived through vision; this provides another basis for deciding that any object, including the body itself, is straight or tilted. Ordinarily we cannot tell whether, at any given side or because of the way we look in relation to the many verticals around us. One of the experimental situations we have developed separates these two main determinants of perception on the position of body in space. It is shown in Figure 1. The apparatus consists of a small room that can be tilted clockwise

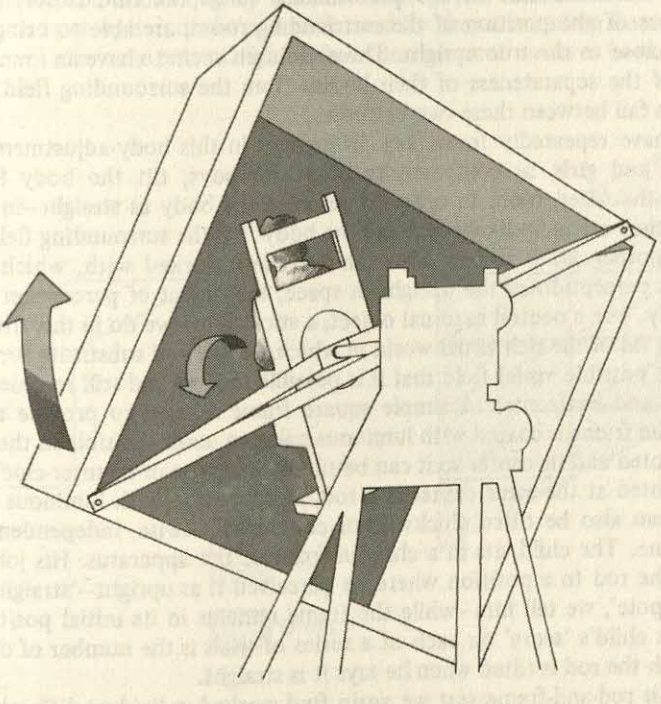


Fig. 1 Apparatus for body-adjustment test

or counter-clockwise; into this room projects a chair that can also be tilted clockwise or counter-clockwise. The child is seated in the chair while room and chair are upright; he then closes his eyes and we tilt the chair and the room to prepared positions. The child then opens his eyes, and if he experiences himself as tilted, the experimenter, standing at the rear of the apparatus, moves the chair at the child's instruction until the child says he is upright. A series of such trials is given. The child's score on each trial is the number of degrees by which the body is tilted when he reports that it is straight. Children find this a marvelous game. If any large group of children or adults is tested in this situation, marked individual differences are found in how they perform the task. Some children, in order to perceive their own bodies as straight, need to have the body aligned with the tilted room. This may sometimes mean being tilted as much as 35 degrees or even more. It may be hard to imagine that anyone can be tilted by that amount and report that he feels straight, that this is the way he sits when he eats his dinner, that he feels no pressure to either side. For children who perform in this way the perception of the position of their own bodies in space seems to be dictated by the relation between the body and the surrounding world. Such a performance suggests a fusion between body and field in experience. At the other extreme of the performance range, we find children who, regardless of the position of the surrounding room, are able to bring their bodies close to the true upright. These children seem to have an immediate sense of the separateness of their bodies from the surrounding field. Most children fall between these two extremes.

We have repeatedly found sex differences in this body-adjustment test. Women and girls, as compared to men and boys, tilt the body farther toward the tilted room in order to perceive the body as straight—in other words, they are more likely to make the body 'fit' the surrounding field.

In another kind of test situation we have worked with, which again involves perception of the upright in space, the object of perception is not the body, but a neutral external object, a stick. What we do in this situation is to get rid of the rich visual world in which we live and substitute for it the simplest possible visual field that it is possible to have and still represent the vertical and horizontal. A simple square frame is used to provide such a field. The frame is coated with luminous paint to make it visible in the dark. It is pivoted and its center so it can be tilted clockwise or counter-clockwise; and pivoted at the same center is a rod, again coated with luminous paint, which can also be tilted clockwise or counter-clockwise, independently of the frame. The child sits in a chair in front of the apparatus. His job is to adjust the rod to a position where he perceived it as upright—'straight with the flagpole', we tell him—while the frame remains in its initial position of tilt. The child's 'score' on each of a series of trials is the number of degrees by which the rod is tilted when he says it is straight.

In this rod-and-frame test we again find marked individual differences in performance. Some children perceive the rod as upright only when it is fully aligned with the tilted frame around it. If the frame is tilted 30 degrees to the left, they tilt the rod 30 degrees to the left; if the frame is tilted 30 degrees to the right, they tilt the rod 30 degrees to the right. 'Straight' for them is defined pretty exclusively by the standard which the frame immediately surrounding the rod provides. There are other children, at the opposite extreme, who are able to deal with the rod more or less independently of the surrounding frame. Their perception of the position of the rod is not dominated by the context in which it occurs. They are able to perceive parts of a field as discrete from the field; in other words, their perception is analytical. Most children fall between the two extremes just described. And again we get differences between boys and girls, and between men and women too. Girls and women require that the rod be tilted farther toward the tilted frame in order to perceive the rod as upright.

What is particularly important for our present discussion is that if the same children are tested in the rod-and-frame test and in the body-adjustment test we considered just before, marked consistency in performance is found. The child who cannot keep his body separate from the surrounding room is also unable to keep the rod separate from the surrounding frame. Or, to put it conversely, the child for whom perception of body position is dictated by the surrounding field is also likely to be the child whose perception of the rod is dictated by the surrounding frame.

Figure 2 shows a pair of figures from another test we have been using.

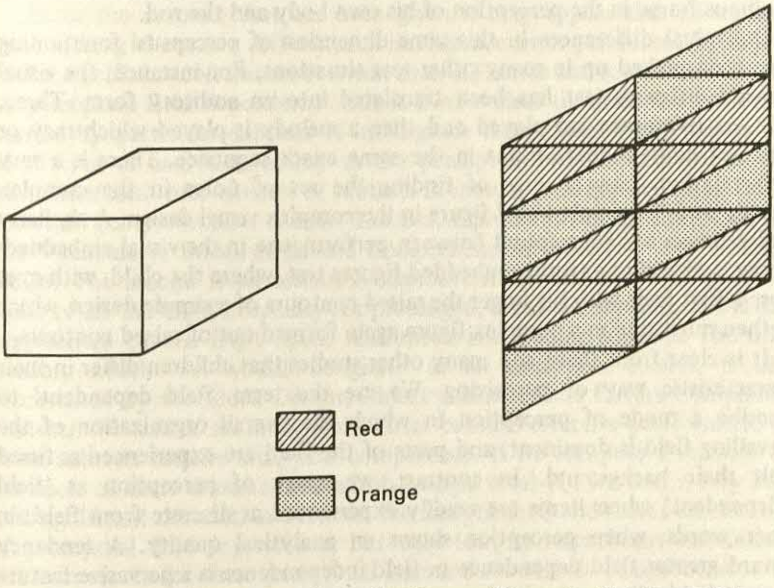


Fig. 2 Sample of simple and complex figures of the embedded-figures test

The child's task here is to locate the simple figure shown on the left in the complex design on the right. In the actual test we do not let the subject see simple and complex figures together. Although this test does not involve orientation toward the upright or the body directly, there is a basic similarity between what is required of the child here and what is required of him in the first two test situations we considered. Here again he is asked to deal perceptually with an item in a field and the question once more is how much does the field influence his perception of the item within it. Specifically, the field (that is, the complex design) is so organized as to embed the simple figure. The measure we obtain of the extent to which perception of the simple figure is influenced by the complex field around it is the amount of time it takes the child to locate the simple figure in the complex design in each of twenty-four such pairs of simple and complex figures. In this test again we find marked individual differences in performance. For some children the simple figure almost 'pops out' of the complex design. Other children are not able to find it in the five minutes allowed. Again most children fall between these two extremes of performance. And once more sex differences are evident: women and girls find it harder to break up the complex configuration and find the simple figure within it.

Of particular importance to us now is the consistency in performance found between this situation and the first two situations. The child who finds it hard to overcome the influence of the complex design in this embedded-figures test is also more influenced by the tilted room and tilted luminous frame in the perception of his own body and the rod.

Individual differences in this same dimension of perceptual functioning have been picked up in many other test situations. For instance, the visual embedded-figures test has been translated into an auditory form. Three, four, or five notes are played and then a melody is played which may or may not contain these notes in the same exact sequence. There is a very high relation between ease of finding the set of notes in the complex melody and the simple visual figure in the complex visual design. A similarly high relation has been found between performance in the visual embedded-figures test and in a tactile embedded-figures test, where the child, with eyes closed, first feels with his finger the raised contours of a simple design, which he then must find in a complex figure again formed out of raised contours.

It is clear from these and many other studies that children differ in their characteristic way of perceiving. We use the term 'field dependent' to describe a mode of perception in which the overall organization of the prevailing field is dominant, and parts of the field are experienced as fused with their background. In contrast, we speak of perception as 'field independent' when items are readily experienced as discrete from field; in other words, when perception shows an analytical quality. A tendency toward greater field dependence or field independence is a pervasive feature of an individual's perceptual functioning.

Individual differences in intellectual functioning

This picture of self-consistency that has been described has now been extended by the demonstration that a tendency to function in a more analytical or less analytical fashion characterizes an individual's intellectual activity as well as his perception. There is now a great deal of evidence that children whose perception is field dependent do less well at solving problems in which essential elements must be isolated from a particular context and recombined into a new relationship. An example of such a problem is one used in a study by Frances Harris. The subject is given a stick and asked to make it fit across an open doorway, but the stick is too short. Subjects quickly learn that to solve the problem they must use a wedge to make the stick stay in place. The experimenter has left such a wedge in the room but he has deliberately 'embedded' it. On the table in the experimental room is a bottle with a stopper, and the stopper is just the right width to serve as the wedge. In order to use it for this purpose, however, it must be taken out of the functional context of stopping up a bottle and used in a very different (and less familiar) functional context of serving as a wedge. There is an extremely high relationship between ability to solve problems of this kind and extent of field dependence. The person who has difficulty with such problems is likely to be very field dependent in our perceptual tests. This should not be surprising, since in both kinds of situations the issue is the person's ability to free himself from the dominant organization of the task material and to function analytically.

From the account that has been given it may appear that all we have been talking about is a difference in intelligence among children. There is in fact a significant correlation between *total* IQ scores and scores for our tests of perceptual field dependence. But it is now clear that this relationship is carried by particular subtests of intelligence tests, those which require the same kind of analytical ability as the perceptual tests themselves. Studies have identified three clusters of subtests in standard intelligence tests. One is a verbal–comprehension cluster. The best representatives of this cluster are the Vocabulary, Information and Comprehension subtests of the Wechsler scales. The second is an attention–concentration cluster, which picks up most of all the child's capacity for prolonged, focused attention, and is best represented by the Digit Span, Arithmetic and Coding subtests. The third cluster, which we would designate as an analytical cluster, is best represented by the Block Design, Object Assembly and Picture Completion subtests. We call it an analytical cluster because effective performance on these subtests require analytical competence. If we compute a separate IQ for each of these clusters, instead of a single total IQ, we find a very high and significant correlation between the analytical IQ and scores on the tests of perceptual field dependence, but only low, non-significant correlations between scores for these perceptual tests and the verbal–comprehension IQ and attention–concentration IQ (Goodenough and Karp, 1961). In other words, if we break up standard intelligence tests into their three main components, what we find is that children who are field independent are superior to field-dependent children only on the analytical cluster of subtests, which require the same kind of analytical ability as the tests of field dependence. They are not identifiably different in their performance on the verbal subtests or the attention-concentration subtests. We may say then that some subtests of standard intelligence tests pick up the same dimension of cognitive functioning as the perceptual tests, and it is this which is responsible for the observed relation between scores for the perceptual tests and total IQ. Certainly, we cannot say that field-independent children are superior in *general* intelligence.

Evidence such as this further confirms the longstanding impression that the IQ, or intelligence, for that matter, cannot be considered a unitary entity. Quite different combinations of subtest scores may go to compose a given IQ; and what we call intelligence may be compounded out of a variety of cognitive patterns. It is therefore necessary to go 'beyond the IQ', and beyond the concept of intelligence as a uni-dimensional affair. The finding that one of the three clusters observed in the standard Wechsler scales connects with a congruent dimension in perceptual functioning provides a broadened basis for thinking about cognitive patterns. We will return to the issue of cognitive patterning later, and consider in particular its significance for the problem of mental retardation.

Individual differences in cognitive (perceptual and intellectual) functioning

To continue with our main story, the studies cited, and numerous other ones as well, have made it quite clear that the style of functioning we first picked up in perception, where we are dealing with an immediately present stimulus configuration, manifests itself as well in intellectual activity, where we are dealing with symbolic functioning. As noted at the outset, we use the designation 'cognitive style' to refer to this kind of characteristic, self-consistent way of functioning that an individual shows across perceptual and intellectual (i.e. cognitive) activities. The particular cognitive style we have been discussing, of which field dependence or field independence is the perceptual component, may be described most broadly as follows: At one extreme there is a tendency for experience to be diffuse and global; the organization of a field as a whole dictates the way in which its parts are experienced. At the other extreme the tendency is for experience to be delineated and structured; parts of a field are experienced as discrete and the field as a whole as structured. To these opposite poles of the cognitive style we have applied the labels 'global' and 'articulated'. It should be emphasized that there is no implication here and the world is populated by two kinds of human beings. Scores for any large group of people on tests of this cognitive style show a continuous distribution, and depending on which side of the mean or average a person's score falls, we say that his cognition is more articulated or more global. It is clear from the evidence on hand that a tendency toward a more global or more articulated mode of functioning pervades a child's cognitive activity; and it may be added, on the basis of other evidence, that a given style of cognitive functioning is a stable characteristic of a child even over very long periods of time. We will return to the question of stability later on.

Individual differences in related personal characteristics

Let us consider other differences we have found between children with a relatively global and relatively articulated mode of cognitive functioning–how they differ 'as people', or in their personalities. We have done a great deal of work on the problem but only a few highlights will be given.

Body concept. One area we have explored with regard to differences in personal functioning is the body concept. 'Body concept' refers to the impression, not necessarily conscious, which the person carries around with him of what his body is like and how it is 'put together'. This impression may be clear and definite or it may be very vague. The image of the body is certainly an important component of the impressions one has of one's self. What we are doing now is to turn the spotlight from experience which has its primary source 'out there', which we have been talking about mainly until now, to experience which has its primary source 'within'. We have found that children who show a more articulated quality in their perceptual and intellectual functioning also give evidence of a more developed–or we may now call it more articulated–body concept, that is, they experience their bodies as having definite limits or boundaries and the parts within as discrete yet interrelated and formed into a definite structure.

Now if we go back to the body-adjustment situation considered at the beginning we can see that from a child's performance in that situation some inference may be drawn about the articulation of his body concept. Take, for example, the child who in order to perceive his body as upright, aligns it with the room tilted at 35 degrees, and in that position reports no feeling of tilt. Such fusion of body and field in experience suggests an absence of body boundaries, psychologically speaking, thus indicating a lack of one of the important characteristics of an articulated body concept.

Another approach we have used in the study of the child's concept of his body is the figure-drawing technique. The child is asked to draw a person and then, after he is finished, to draw a person of the opposite sex. The idea behind this technique is that in sketching the body on paper the person is dealing with the body vicariously and in doing so projects something of his view of what the body is like, a view compounded out of experiences he has had with his own body and the bodies of others. Before discussing the method we use to evaluate figure drawings for evidence of articulation of body concept, it would be useful to examine drawings (Figures 3 and 4) made by children whose performance in cognitive tests placed them at the

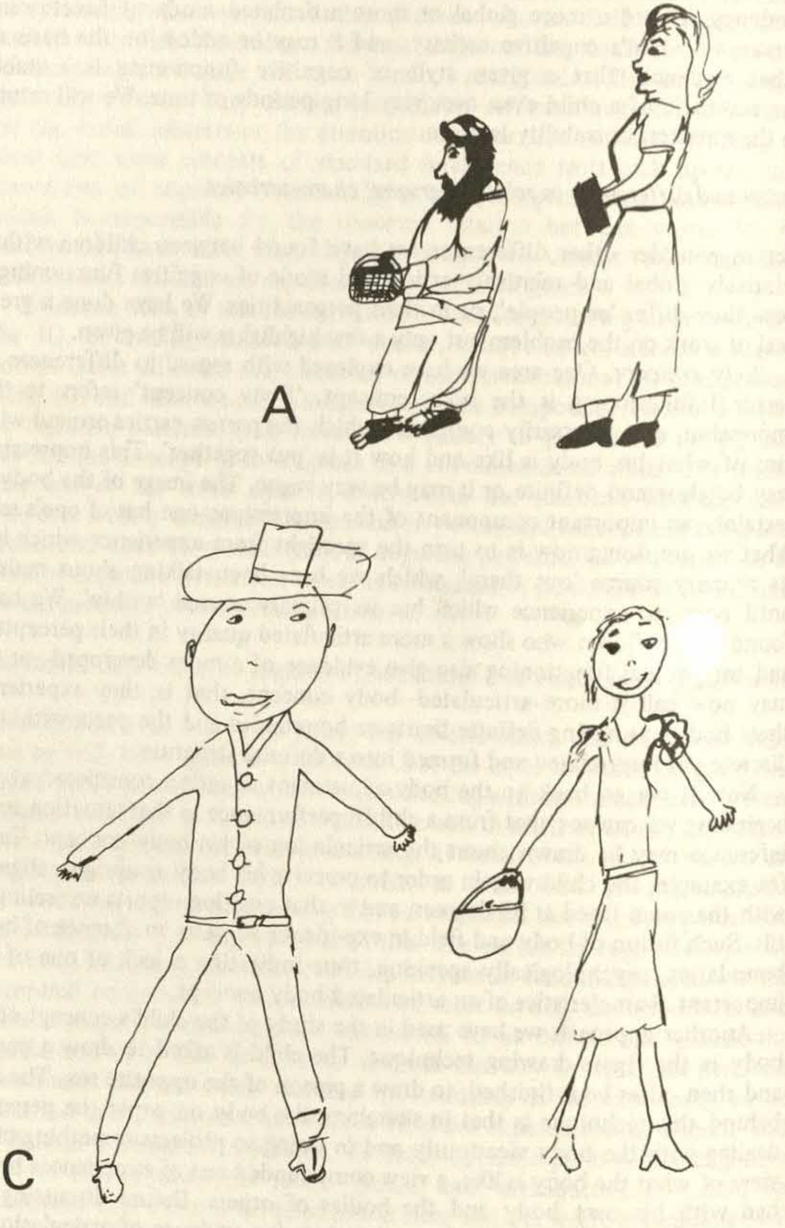


Fig. 3 The four pairs of drawings were made by four ten-year-old boys whose perception was field independent

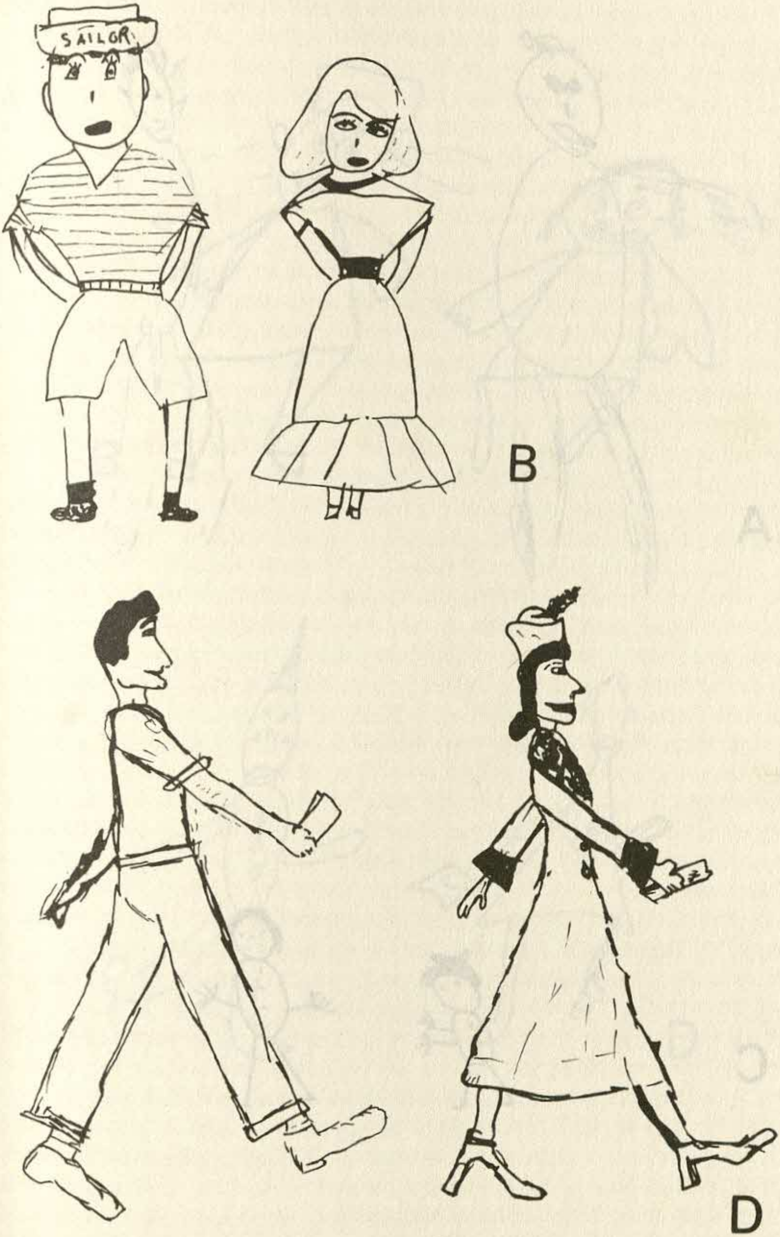


Fig. 3 (continued) **In each pair, the figure on the left was drawn first**

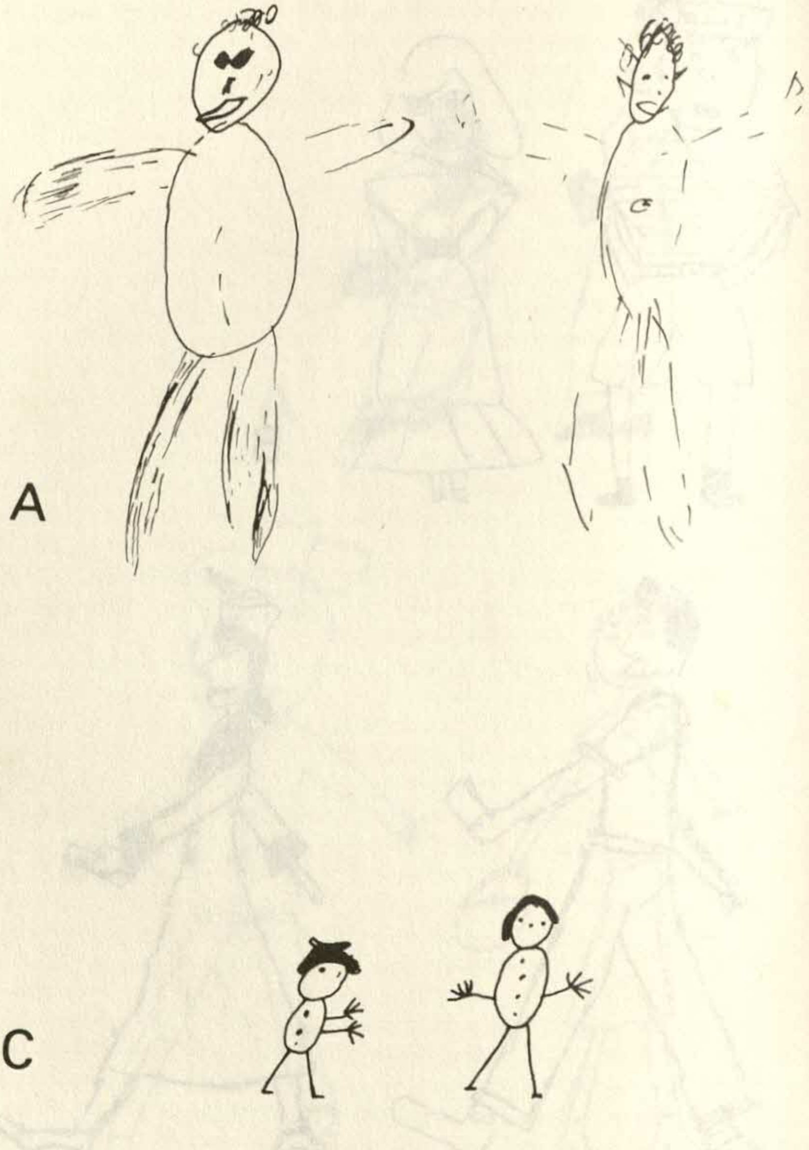


Fig. 4 The four pairs of drawings were made by four ten-year-old boys whose perception was field dependent.

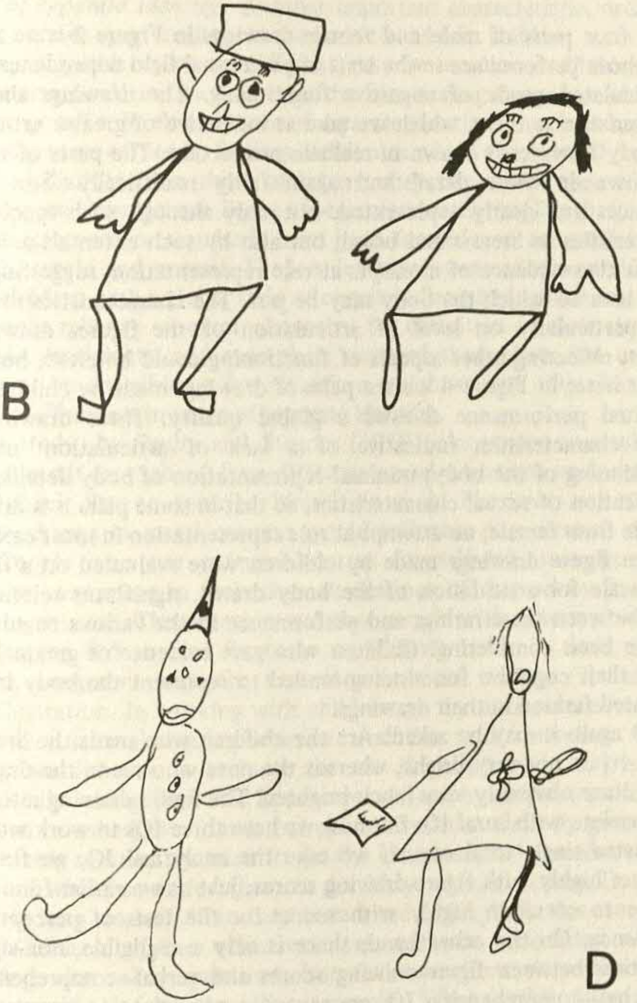


Fig. 4 (continued) **In each pair, the figure on the left was drawn first.**

extremes of the articulation–global dimension. The sample drawings shown are all made by ten-year-old boys recruited from a neighboring elementary school.

The four pairs of male and female drawings in Figure 3 were made by boys whose performance in the tests of perceptual field dependence showed an articulated mode of cognitive functioning. The drawings show such characteristics as these, which we take as indicative of greater articulation: The body is typically drawn in realistic proportion. The parts of the body are shown in some detail and again fairly realistically. Sex and sex differences are clearly represented, not only through such specific body characteristics as breasts and beard, but also by such externals as clothing. There is also evidence of attempts at role representation, suggesting a sense of the uses to which the body may be put. The characteristics mentioned speak particularly on level of articulation of the figures drawn; other features, reflecting other aspects of functioning could be cited, but are not relevant here. In Figure 4 we see pairs of drawings made by children whose perceptual performance showed a global quality. These drawings show various characteristics indicative of a lack of articulation: unrealistic proportioning of the body; minimal representation of body details; little or no indication of sexual characteristics, so that in some pairs it is difficult to tell male from female; no attempt at role representation in most cases.

When figure drawings made by children were evaluated on a five-point rating scale for articulation of the body drawn, significant relations were found between these ratings and performance in the various cognitive tests we have been considering. Children who gave evidence of greater articulation in their cognitive functioning tended to represent the body in a more articulated fashion in their drawings.

Now again it may be asked: Are the children who made the drawings in Figure 4 just not very bright, whereas the ones who made the drawings in Figure 3 are obviously very much brighter? The figure-drawing ratings do in fact correlate with *total* IQ. But now we have three IQs to work with rather than just a single total one. If we take the analytical IQ, we find that it correlates highly with figure-drawing scores, just as we earlier found this IQ measure to correlate highly with scores for the tests of perceptual field dependence. On the other hand, there is only a negligible, non-significant correlation between figure-drawing scores and verbal–comprehension IQ; and verbal–comprehension IQ, we saw, did not relate to perceptual tests scores either. In other words, the children who made the drawings in Figure 3 and those who made the drawings in Figure 4 could not be told apart in their performance on the three subtests of standard intelligence tests which go to make up the verbal–comprehension IQ (Vocabulary, Information and Comprehension). They are markedly different on Object Assembly, Picture Completion and Block Design, the subtests used to determine the analytical IQ.

From the evidence presented and other findings as well we can say that children with a more articulated or more global mode of cognitive functioning are different in the nature of their body concept.

Sense of separate identity. Another important characteristic, ordinarily subsumed under personality, in regard to which children with these contrasting cognitive styles are different, may be mentioned. Children who show greater articulation in their cognitive functioning also give evidence of what we may call a developed sense of separate identity. That is to say, they have an awareness of needs, feelings, attributes which they identify as their own and as distinct from those of others; they experience themselves as persons separate from others. Sense of separate identity implies that internal frames of reference or standards have been formed and are available for viewing the world and oneself. The less developed sense of separate identity found in children with a global style of cognitive functioning manifests itself in continuous reliance on external sources for definition of their attitudes, judgements, sentiments, and even their views of themselves. Let us make this concrete by describing a few situations which we and others have used to identify differences in sense of separate identity.

In one study, by Rudin and Stagner (1958), the person was asked to think of himself in each of four different social situations and, using an adjective check list, to describe himself in each. The issue was to what extent the change in social context would influence the person's view of himself. People who use external sources for self-definition would be expected to vary more in their characterizations of themselves from one situation to another. The outcome was that people with a more global way of experiencing did in fact fluctuate significantly more in their view of themselves, congruent with the variations in social context. Let us take another illustration. In working with children we try to observe as much as we can of their natural ongoing behavior in relation to the examiner. One of the things we did was to evaluate the use the child made of the adult examiner while taking Thematic Apperception Test, by examining the kinds of questions he asked. We found that children with a relatively global way of experiencing more often tried to get the examiner to define the task for them, to tell what he wanted them to do. They asked questions like: 'Do I have to tell long stories?' 'Can they have a sad ending?' 'Is this story long enough?' Children at the other extreme, with an articulated way of experiencing, also asked questions, but their questions were directed at obtaining specific information, for example, 'What is this thing in the picture here?' They seemed to have worked out their own standards for dealing with the task of making up stories, and therefore did not require guidance from the examiner.

We may express these findings about sense of separate identity by saying that the child for whom the frame around the rod is very influential in determining how the rod is experienced, is similarly strongly influenced by the immediate social context in his experience of himself.

It may be stated, in overview, that the contrasting constellations of characteristics we have observed in children seem to reflect a developmental difference. The difference is in the extent of differentiation they have achieved. We consider it less differentiated if the child's experience of the world is global. We consider it less differentiated if his image of his body is like a blob, with the parts not discrete. We consider it less differentiated if he does not have a ready feeling of himself as a person apart from others. We consider it less differentiated if he uses global defenses such as repression and denial.

It must be stressed emphatically that, though we seem to be dealing with differences in development of differentiation, there is no implication of differences in adjustment or extent of pathology. Children at the global extreme are as likely to show pathology as children at the articulated extreme. In fact, we have the impression that there is more pathology among children at the extremes than in the middle range. Our idea is that both a high level of differentiation and limited differentiation can occur in a setting of either effective or impaired integration. What is of interest is that there are differences in the *kinds* of pathology that tend to develop at the two extremes when disturbances in behavior occur. For example, among adults we are likely to find obsessive-compulsives at the articulated end and hysterics at the global end. Among children, it is at the global end that we find youngsters with impulse-control problems. It is at the other end that we get emotionally isolated children, over-intellectual children, children with strong perfectionist tendencies.

This very sketchy overview has sought to demonstrate that cognitive functioning is intimately related to personality. In fact it is sometimes hard to tell where cognition ends and personality begins. It is also worth mentioning that to the extent that we can pick up in the child's cognitive activity a rather deep and important dimension of his general psychological functioning, and to the extent that we can study cognition in the laboratory, we have in cognition an experimental route to the study of the child's overall psychological organization. We are inclined to think of cognitive performance, because of its objective character, as providing a 'tracer element' for pegging the child's status in important psychological domains.

Origins of cognitive style

Now there is the important question of how the differences among children that we have been describing arise. The cross-sectional and longitudinal studies we have done leave little doubt that during development the cognitive functioning of children shows, in general, a progression from global to articulated. It is also indicated, however, that as early as the age of five children already differ in the extent to which their experience tends to be articulated or global; and in longitudinal studies, where the same children were followed over a period of years, marked stability of functioning was clearly evident. This stability was indicated by the high test-retest correlations of scores on our tests. As one example, when the figure drawings obtained from a group of boys at age ten and again at age twenty-four were rated for articulation of body concept, the ratings at the two ages, fourteen years apart and covering an important period of growth, correlated 0.81.

There is no contradiction between saying, on the one hand, that children as a group move toward greater articulation of experience as they grow older, and saying, on the other hand, that there is high stability in children's functioning in this dimension over time. Another way of stating these findings is that children tend to hold their relative positions in their group during development. Thus, the child who at a very early age gives evidence of a global way of experiencing, relative to his peers, will at later ages again show a global way of experiencing, relative to his peers, though he and his peers as a group, have changed in the direction of greater articulation.

These findings together suggest that we are dealing with individual differences in *development of differentiation,* so that the question raised a moment ago may now be put more specifically by asking what are the sources of differences among children in development of the early-established, stable, deep-seated clusters of characteristics that we subsume under differentiation? We have done two kinds of studies as first steps toward answering this question (Dyk and Wilkin, 1965; Goodenough and Karp, 1961). First, we have studied patterns of relation to the family, and particularly to the mother, of children who differed in cognitive style, or to put it more generally, since this was our concept at the time, who differed in level of differentiation. Second, we have carried out longitudinal studies with two groups of children. In one case, already mentioned, we followed a group of boys from age ten to age twenty-four; in a second case we restudied at ages six through nine a group of children studied by Escalona and her collaborators in early infancy. Since the analyses of data from these longitudinal studies are still under way, it is not possible to say much about them.

Our main study of mother–child relations, in connection with the development of differentiation in children, was cross-sectional in nature and made use of a group of ten-year-old boys and their mothers. The most reasonable approach to the study of mother–child relations is an interactional one. We can no longer take the view that, on the one hand, the infant is an empty container into which the mother pours whatever psychological content she wishes, or, on the other hand, that the mother's behavior is exclusively a reaction to the kind of child she brought back from the hospital with her. Effective application of an interaction approach requires a longitudinal study, in which the investigator is on the scene to observe the mother even before the baby is born, then studies the baby when he appears, and after that observes continuously the growing relation between them. For practical reasons we could not do such a study at the time we began this line of investigation; we started with cross-sectional studies, supplemented by the two longitudinal studies mentioned earlier, thinking that these cross-sectional studies would suggest hypotheses to be tested in the longitudinal studies later on. Our exploration of mother–child interactions sought to obtain evidence about these interactions both from the mother's standpoint, through an interview with her, and from the child's standpoint, mainly through his fantasies about parental role as projected in TAT (Thematic Apperception Test) stories he made up concerning mothers and their children.

The interviews with the mothers of ten-year-old boys, conducted in the home by a staff member who knew nothing about the child, examined the mother–child relation in the immediate present and in historical perspective as well. On the basis of the impact of the total evidence gained from the interview the mothers were classified into two groups. In one group were placed mothers who gave the impression of having interacted with their children in a way which permitted, fostered, or even pushed the child's progress toward differentiation. In the second group were placed mothers who interacted with their children in a way that seemed to have hampered the child's progress toward differentiation. Although the interviews were 'freewheeling', they focused, first of all, on the opportunity the mother gave the child to separate from her, and, second, on the mother's role in imparting standards or frames of reference for internalization–in other words, what the mother did to help the child develop what was earlier called a sense of separate identity.

Let us make more concrete what went into these two ratings by enumerating some of the specific clues used in judging, for instance, that a mother had interfered with her child's progress towards differentiation. One was that because of fears and anxieties about her child, or especially strong ties to him, the mother restricted the child's activities markedly. We had one mother, for example, who took her son out of the Boy Scouts at age ten because they were too rough. A second mother escorted her child to school, just a few blocks away, something very few mothers in the group studied did for boys of ten years old.

Another clue was that the mother's behavior was not in the direction of helping the child achieve increasing responsibility and self-sufficiency. Her handling of the child's allowance was particularly informative here. Some mothers gave their ten-year-old sons an allowance and, however much it pained them, let the child spend it in any way he wanted. Other mothers gave their children no allowance at all. One son of such a mother, when asked about his allowance, said 'Who, me, I'd only spend it'. Another clue was that the mother's physical care of her child seemed inappropriate to his age. We had mothers of these ten-year-old boys who still bathed their sons, and helped them dress. Still another clue was that the mother regarded the child as delicate, for instance as too weak to perform chores about the house or to engage in certain kinds of sports. These are examples of the kinds of specific items of behavior considered in judging that a mother had interacted with her child in a manner likely to interfere with his differentiation.

Obviously a mother could end up in this group through a thousand and one different pathways of interaction with her child. Let me stress too that the classification made was not one of 'good' and 'bad' mothering. We had one mother, for example, judged to have fostered differentiation in her son, who gave him a dollar for every point over 95 he received at the end of the semester in each of his subjects. That is quite a start in life. The child, in our clinical studies of him, showed the adverse effects of such extreme pressure upon him. He was a quite depressed child who again and again gave evidence of feeling that people expected more of him than he could deliver.

The two-fold classification made of mothers, on the basis of the interviews with them, agreed significantly, for the most part, with measures of differentiation for their children. That is to say, children of mothers judged to have fostered differentiation gave evidence of field-independent perception and, in general, an articulated cognitive style; an articulated body concept; a developed sense of separate identity; and use of developed, specialized defenses.

These relationships were essentially confirmed when mother–child interactions were explored from the child's point of view. The children were asked to make up stories in response to TAT pictures, and we picked out for analysis those stories that dealt with relations between parents and their children. We undertook this analysis with the idea that mothers who interacted with their children in the contrasting ways we discerned in our interview with mothers, would be experienced differently by their children, particularly as more supportive or less supportive. And in fact this was the outcome. Children with a more global way of experiencing and with other characteristics of limited differentiation tended to cast parents in the less supportive role; to characterize parents as irrational in their treatment of their children; and to describe parents as less able to give their children help when they needed it.

It is interesting, methodologically, that the ratings made of mothers on the basis of interviews with them were in significant agreement with the ratings of the children's parent–child TAT stories. Children's projections about parental role in TAT stories thus have something to do with what the parents are really like when seen in the flesh.

It is perhaps evident even from this very brief description of our studies that the intellectual and perceptual functioning of the child must be viewed as embedded in a personal life and in a personal history.

Implications for problems of education

The account that has been given of one rather intensively studied cognitive style was intended to convey some idea of what cognitive styles are like. Using this particular cognitive style as a reference let us consider now the ways in which knowledge about the existence of cognitive styles and their specific nature may be useful in dealing with some of the problems encountered in the educational situation. In particular, let us examine the implications of the cognitive-style work for issues of evaluation, placement and teaching methods.

Evaluation

With regard to evaluation, it seems quite evident that children may be effectively characterized in terms of cognitive style; and such characterizations are particularly appropriate to the kind of 'sizing up' of children needed in the educational situation. Cognitive style encompasses the perceptual as well as the intellectual functioning of a child, and in this regard covers cognitive activity more comprehensively than our usual intelligence tests, which emphasize intellectual functioning. A given cognitive style also manifests itself pervasively throughout a person's cognitive activity. Moreover, as we have seen, a cognitive style appears to be the manifestation, in the cognitive sphere, of a broader dimension that extends into the domain of personal functioning–or personality–and so carries a message about the child as a person. In fact, performance in our tests of perceptual field dependence is coming to be used as a source of information about personality. In this connection, it was suggested earlier that perceptual performance may be taken as a 'tracer element', identifying a large psychological dimension of which such performance is a part. Thus, when we speak of a child's cognitive style we are covering a great deal of psychological territory in our characterization of him.

The cognitive-style approach to evaluation may be used to supplement the assessment provided by intelligence tests. In fact, it is not far-fetched to imagine that test batteries emerging from cognitive style research may in time replace intelligence tests. Historically, the development of intelligence tests outran the development of a theory of intelligence. This is understandable, since under the compelling need to classify children in the schools, test development could not await adequate theory. The tests that have emerged accordingly have more of an empirical basis than a theoretical basis. Essentially they consist of groups of tasks which, by experience, have been found to discriminate between slow-learning and fast-learning children, that is, children apparently differing in intelligence. But we do not know as much as we should about the psychological process involved in carrying out these tasks. Nor do we always have an adequate rationale for using the particular tasks we do. In fact, there is the paradox that much work has been going into finding a conceptual rationale for these tasks *after* they have long been in use.

It is possible that the work on cognitive style is following the course toward test development that might have been taken in the first place (and most appropriately should have been taken), had time allowed. This is a progression from theory, to specific test rationale, to test construction. As an illustration, each of the tests in the battery we used to pin-point the articulation–global dimension of cognitive functioning has a specific rationale derived from a particular conception of the dimension.

Obviously, the definition of a child's cognitive make-up requires consideration of a variety of cognitive characteristics. What is needed is a 'cognitive map', as it were, of the individual child, reflecting the pattern into which his cognitive characteristics fall. Cognitive styles may provide the main ingredients out of which these maps are composed. A great deal more must be known about cognitive styles than we do now before they can serve in the task of mapping. It is, however, already possible to see how much more useful it is for understanding a child's cognitive functioning to consider his cognitive patterning rather than just his IQ. Let us take a few examples.

We considered earlier the three-way breakdown of standard intelligence tests into verbal-comprehension, attention–concentration and analytical components; and, as we saw, measures of perceptual field dependence relate significantly to the analytical component only. In the course of our studies we have encountered children who show marked discrepancies in their level of performance in these various areas. Two boys come particularly to mind. Both had quite high total IQs (in the 120's), but behind these IQs was a strikingly discrepant pattern. These boys performed at an extremely high level on the three tests we use to compute the verbal–comprehension IQ (Vocabulary, Information and Comprehension) and at relatively low level on the analytical IQ sub-tests (Object Assembly, Picture Completion and Block Design). For one of them, for example, the summed scaled scores for the verbal-comprehension triad was 52 and for the analytical triad only 25 (30 is the average score for the general population). Corresponding to their low level of performance on the analytical sub-tests, both boys were markedly field dependent. Moreover, an evaluation for them in other areas of functioning gave evidence of limited differentiation. For reasons we do not yet know, these boys have apparently invested heavily in verbal activities, without corresponding investment in other areas of functioning. They are children who are more comfortable 'talking about things' than dealing directly with real situations. They use their verbal facility to good advantage in relation to others. We found that many of us, in the notes we recorded for each of these boys after the test sessions with him, commented on what a good 'conversationalist' he was. That he was using his conversation to impress became clearly evident, however, when he somehow managed to work in the fact that he regularly read 'obstetric and geneology' books which he picked up in the trash can outside the house of an obstetrician who lived across the street. We found this boy comfortable with adults but unable to get on with other boys, whose company he actually seemed to avoid. This is understandable since boys require more than good conversational ability to get along with their peers. Altogether, the personality structure of these boys, as judged from projective tests, was far less developed than might be imagined from talking to them.

The characterization of the boys in terms of high verbal competence and limitedly developed articulation is far more informative than a statement about their IQs. Moreover, such a characterization obviously encompasses the domain of their personal functioning.

With these two boys, it would obviously not get us very far to ask how intelligent they are. To answer this question we would have to specify which kind of intelligence is meant. It is interesting that the abbreviated intelligence tests commonly used for a quick assessment of a child's intelligence are often verbal; sometimes in fact a vocabulary test is used. Just imagine how we would be misled in judging these boys if one of these 'quickie' tests were used.

Another illustration of the usefulness of a cognitive-mapping approach to an understanding of cognitive functioning is found in work we have done with retarded children (Witkin *et al.*, 1966). Among retarded children in the 60-80 IQ range we found that, though in general they are lower then normal children in all the cognitive areas we have been considering, they fell below much more in verbal-comprehension IQ than in analytical IQ. Among such children measures of perceptual field dependence again relate significantly to analytical but not to verbal IQs. Retarded children thus show a discrepancy in the opposite direction from that observed in the two boys just discussed. This discrepancy has been found very consistently in retarded children we have studied thus far. It was found among boys from Letchworth Village, a state institution for the retarded, as well as boys living at home and attending special classes for the retarded (CRMD classes) in neighborhood elementary schools. The discrepancy thus cannot be accounted for in terms of a particularly adverse effect of institutional living on verbal functioning. Not only was the discrepancy consistently observed in individual cases, but it was quite large in magnitude. For example, for the Letchworth group the mean summed scaled score for the triad of tests going to the analytical IQ was 20.1, the corresponding value for the verbal–comprehension triad was 11.1. (In each case 30 is the average for the general population.) In some individual cases, we found analytical IQs of over 100, with an extremely low verbal-comprehension IQ.

Considering IQ alone obscures the important fact that retarded children are not uniformly impaired in all areas of cognitive functioning. The resources that at least some of these children seem to have may be used to good advantage in helping them work out a life adjustment.

It was a surprise to us that though the retarded groups we worked with were unselected, the cognitive discrepancy almost always went in the direction described. From the way in which intelligence tests are made up, we might reasonably have expected cases with an opposite discrepancy to occur just about as frequently. Our belief is that this peculiar outcome may be a function of the heavy positive emphasis we place on verbal competence in routing children through life, to the relative disregard of our other kinds of ability. A child with cognitive impairment is especially likely to be picked up if his difficulty is in the verbal area. Because this difficulty is likely to show itself very readily, he will easily be identified by his parents as 'not quite all right'. In school, where the emphasis is strongly on verbal abilities, as speech, reading, etc., his difficulty will also be quickly noted. Such a child is much more likely to be sent for special testing, and to end up in a class for children with retarded mental development (CRMD), than the low IQ child with relatively better verbal competence and with impairment in the analytical area or in articulation more generally. The latter kind of child may not only escape the selection filter, which leads a low IQ child toward CRMD classes and therefore identification as retarded, but he may even be considered 'good' by his teachers. For example, we know from our studies of field-dependent children that they are often likely to conform to adult authority. This quality, together with their ability to 'talk nicely', is apt to be particularly pleasing to their teachers. We believe such children exist in our schools, but because large-scale intelligence testing of children other than those presenting intellectual or emotional difficulties is not done, we do not have evidence to support this view. We may further imagine that because of the cognitive deficit they suffer, and their limited differentiation, these children are likely to get into psychological difficulty, though perhaps in other ways than the retarded children usually finding their way into CRMD classes.

The possibilities that have been raised may be put to empirical test. But they arise as issues for research only when we go beyond the IQ to an exploration of cognitive patterns.

In these illustrations of the kinds of cognitive patterns that may be found among children, we have considered just a few of the cognitive characteristics out of which these patterns may be compounded as work goes forward on identifying cognitive styles and determining their nature, we will be able to extend the range of characteristics to be considered in composing cognitive maps. Such work is in fact progressing in a number of research centers. Among other cognitive styles investigated in this work are, for example, 'levelling–sharpening', studied by a group of investigators at New York University and the Menninger Foundation, and 'tolerance of ambiguity' studied particularly by Frenkel-Brunswik.

The cognitive maps that we will in time be able to establish for individual children offer promise of providing a rich, sensitive, complex and comprehensive way of characterizing children both in their cognitive functioning and their functioning as persons. Whereas unidimensional assessments like the IQ encourage us to think in terms of 'more or less' and 'better or worse', the cognitive-map concept focuses on the ways in which children are unique and the multiplicity of ways in which they may be different. The result is to emphasize individuality and to de-emphasize labelling along a simple 'better-or-worse' continuum.

Placement

What we have been saying about the implications of cognitive style for evaluation of children is also relevant to the problem of school placement, the second aspect of the educational situation to be considered. To the extent that we do a better job of sizing up children, we have more information on which to base decisions about grouping them in the classroom. Knowledge of the cognitive maps, based on a careful study of well-understood cognitive styles, may provide a sensitive basis for placement. Considering the articulation–global dimension as an example, it is probably not difficult for educators to see how knowledge of children's positions on this dimension would be useful, along with other information about them, in judging how to group them. Obviously, decisions about grouping are complex, as we can see immediately when we raise the question whether children should be grouped on the basis of similarity in cognitive patterns or whether they should be mixed.

The existence of sex differences in cognitive functioning is also important for the placement problem. The tendency for girls to be more field dependent than boys is widespread. This sex difference has now been observed in a large number of studies, using groups differing in socio-economic background and national origins. Sex differences in extent of field dependence have been observed not only in the United States, but in a number of Western European countries and in Israel, Hong Kong and Sierra Leone, Africa, as well. (See Witkin *et al.*, (1962) for review of the evidence on sex differences in field dependence.) Though the differences between the sexes is small, it is a persistent one. Paralleling the difference in analytical competence in favor of boys, it has consistently been found that girls are superior to boys in verbal tasks. How may these differences in cognitive patterns between boys and girls best be taken into account in class grouping? Obviously this cognitive difference is not the only factor important for deciding whether boys and girls should be placed together or apart, or in what ways they should be mixed, but knowing about this factor is surely helpful to a decision.

It is not possible to do justice to the question of placement without emphasizing that we place children not only with their peers, but also with particular teachers. Might not particular combinations of cognitive characteristics of teachers and their pupils be most conducive to effective learning by these pupils? Relevant here is a study which showed that psychiatrists who tested extremely field independent on our perceptual tests tended to favor non-involving types of therapies with their patients, whereas relatively more (though not extremely) field-dependent therapists favored therapy focusing on interpersonal relations between therapist and patient (Pollack and Kiev, 1963). If the therapist's approach to his patient is related to the therapist's cognitive style, may we not expect the teacher's approach to his students to relate to the teacher's cognitive style? And, to complicate matters further, might not the congruence or lack of congruence between cognitive patterns of teacher and student be important for how teaching and learning proceed?

Methods of teaching

Let us turn finally to the question of methods of teaching. It is possible to do no more than throw out some very general suggestions for research into the relation between cognitive style and learning. As we have seen, the articulation–global dimension refers to how children spontaneously experience the world about them and themselves. Our evidence suggests that a tendency toward a more articulated or global style colours the child's experience pervasively from moment to moment. Is it not reasonable to expect that knowing a child's particular cognitive styles may help us determine the best way of reaching him? Determining the methods of teaching that are most effective for children with particular cognitive characteristics is a promising line of future research. The more we understand and are able to identify differences in cognitive patterns of children, the more will we be able to do to devise appropriate teaching methods. To the extent that children are different, methods of teaching them must be different.

As suggested earlier, establishing cognitive maps is a way of identifying resources and weak points in a child's cognitive makeup. This, in turn, can help determine how he may best be taught.

Because cognitive style speaks on personal characteristics, we may expect children with different cognitive styles to differ in their relation to their fellow students and teachers. As one example, we may anticipate, from our findings that children who are field independent and show an articulated cognitive style will be less conforming and more self-directed in their classroom behavior. The way in which personal characteristics associated with cognitive style may intermingle with more purely cognitive characteristics to influence an individual's behavior is illustrated by an area that has been more extensively studied. This is the relation between occupational choice and cognitive style. It has been found that persons with an articulated cognitive style tend to choose occupations which not only draw on their special cognitive skills, but which are unconventional and which do not particularly involve them in relations with other persons.

Still in the area of teaching methods, we may ask whether the teacher can supplement the family's role of fostering the development of differentiation in the child, including the achievement of articulated cognitive functioning. For leads as to how this may be done we can draw on some of our findings on how mothers of more differentiated and less differentiated children behave toward them. Particularly relevant is the mother's behavior as interpreter of the world to the child. We found, for example, that some of the mothers of more field-dependent children, in conveying the dangers to be encountered in this world, represented the world as dangerous *in toto*. An interview with one of these mothers was filled with accounts of the awful things that had happened to children she knew in this danger-ridden world. In contrast to this global concept, mothers of field-independent boys were more likely to specify sources of dangers selectively–for example, the child might fall and hurt himself by climbing onto a weak branch high up on a tree. As another example, some mothers dealt with their children's needs by a universally applied–that is, global–method. One such mother, when her child was an infant, stuck her breast in the child's mouth every time he showed discomfort. Other mothers diversified their methods of satisfying their children's needs, out of an apparent recognition of the diversity of the needs themselves. Although we have not made systematic observations on this point, common experience suggests that whereas some mothers go out of their way to introduce their children to new situations and events by careful description and explanation, other mothers fail to do this, or if they try, are not able to go beyond a very vague account. We may imagine that these two approaches could have quite different consequences for progress toward articulation of experience to the children. Translating these ideas into a general form applicable to the classroom situation, we might imagine that emphasis in teaching on both specificity and structuring (that is, principles) would be most conducive to the development of articulation. Obviously, ways need to be found of teaching particular content areas with an eye to achieving this aim.

Some hope that such teaching methods may prove effective comes from the observation that even blind children, so severely handicapped in ways that matter for articulation, are capable of developing articulated cognitive functioning. Vision, lacking in the blind child, is the sensory channel most useful in the development of articulation. Because through vision we are unable to apprehend parts of the field and the field as a whole in one view, our experiencing of items as discrete from background and the field as structured–in other words, articulated perception–is facilitated. Vision similarly facilitates apprehension of the body itself both as an entity apart from the field and as structured. Moreover, through vision, we can form an impression of how the bodies of others are 'put together', and this in turn contributes to development of an articulated body concept. To arrive at an impression of objects as discrete and of field or body as structured through the senses that remain after vision is removed, is very difficult. For example, with touch or sound, organization can be achieved only by the serial experience of component parts. To compose an impression of structure in this way is so much more difficult than when the field and its parts can be seen at once together. Again, forming an impression of the parts of the body and their interrelatedness through postural sensations alone is not easy. In all these ways, the absence of vision may be expected to hamper the usual progression in cognitive development from global to more articulated. In still another way development of articulation is hindered in the blind child. Because he cannot see, dependence on others is intensified and prolonged. As noted earlier, children who are prevented from separating from their mothers are likely to show a global style of experiencing and limited differentiation.

To test blind children for articulation, our usual tests which require vision could of course not be used. But as mentioned earlier there now exist tactile and auditory embedded-figures tests, and it was possible to transpose some of the other tests we use into a non-visual form. For example, the patterns on the blocks of the Block Designs subtest of the Wechsler were made up of surfaces of different texture, instead of color, as in the original. Studies with congenitally totally blind children (Witkin *et al.*, 1968) have shown, first, individual consistency in performance on tests of articulation, similar to that found in the sighted; and, second, marked individual differences in performance among children. To come back to the starting point of this discussion, that even some blind children can overcome the serious handicaps they suffer and achieve an articulated cognitive style, makes it hopeful that we may be able to devise ways of teaching less handicapped children to move toward greater articulation. Studies of the development of blind children who show this achievement may help us identify effective teaching methods for normal children.

Finally, the comments made earlier about sex differences in cognition bear on methods of teaching. Based in part on the sex differences we found, Waetjen (1962) writing in a recent issue of the *NEA Journal*, has asked, 'Is Learning Sexless?' In this article he comments as follows: 'We can, and should, consider sex differentials in learning. We must try to find out if, in our teaching and general school operation, we have created an official mold for learning. Is it possible that such a mold–albeit intangible and unacknowledged–is shaping our schools and our students? And is this mold one into which girls fit more easily than do boys? No matter what we would like to believe, the truth is that we do not know the answers'. At another point Waetjen offers a specific suggestion: 'Since boys are more field independent and act on internal cues in their perception, it might be worthwhile to place boys in classrooms with teachers who feel comfortable about allowing pupils to make many of their own decisions, to voice differing opinions, and to pursue at length portions of the material to be learned'.

Maccoby (1963), in a paper given at a symposium on 'The Potential of Women', has expressed concern over the fact that in our schools we in effect train women to be field dependent and to develop the other characteristics associated with it, but yet, increasingly, society expects women to assume roles that require a field-independent way of functioning. The result is a conflict which provokes anxiety and hampers the working out of a satisfying role in life. Maccoby writes: 'Suppose a girl does succeed in maintaining, throughout her childhood years, the qualities of dominance, independence and active striving that appear to be requisities for good analytic thinking. In so doing, she is defying the conventions concerning what is appropriate behavior for her sex. She may do that successfully, in many ways, but I suggest that it is a rare intellectual woman who will not have paid a price for it: a price in anxiety'. She then goes on to say: 'Could we not accept and encourage the active, dominant, independent qualities of the intellectual girl without labeling her as masculine, and encourage in her whatever aspects of femininity *are* compatible with an analytic quality of mind?' This is a probing question, loaded with challenge for all of us–parents, teachers and even just members of society who play any part in helping boys and girls grow up.

These are some of the ways in which the work on cognitive style may have value for problems of education. It must be left to educators to assess the usefulness of these suggestions.

References

DYKE, R. B. and WITKIN, H. A. (1965) 'Family experiences related to the development of differentiation in children', *Child Develop., 30,* 21–55.

GOODENOUGH, D. R. and KARP, S. A. (1961) 'Field dependence and intellectual functioning', *J. abnorm. soc. Psychol., 63,* 241–6.

MACCOBY, E. (1963) 'Woman's intellect'. In FARBER, S. M. and WILSON, R. H. L. (Eds.) *The Potential of Woman.* New York: McGraw-Hill, pp. 24–39.

POLLACK, I. W. and KIEV, A. (1963) 'Spatial orientation and psychotherapy: an experimental study of perception', *J. nerv. ment. Dis., 137,* 93–7.

RUDIN, S. A. and STAGNER, R. (1958) 'Figure-ground phenomena in the perception of physical and social stimuli', *J. Psychol., 45,* 213–25.

WAETJEN, W. B. (1962) 'Is learning sexless?', *NEA Journal, 51,* 12–14.

WITKIN, H. A., BIRNBAUM, J., LOMONACO, S., LEHR, S., and HERMAN, J. L. (1968) 'Cognitive patterning in congenitally, totally blind children', *Child Develop., 39,* 767–86.

WITKIN, H. A., DYK, R. B., FATERSON, H. F., GOODENOUGH, D. R. and KARP, S. A. (1962) Psychological Differentiation. New York: John Wiley.

WITKIN, H. A., FATERSON, H. F., GOODENOUGH, D. R. and BIRNBAUM, J. (1966) 'Cognitive patterning in mildly retarded boys', *Child Develop., 37,* 301-16.

WITKIN, H. A., LEWIS, H. B., HERTZMAN, M., MACHOVER, K., MEISSNER, P. B., and WAPNER, S. (1954) *Personality Through Perception.* New York: Harper.

Reprinted by permission of Grune & Stratton, Inc., and the author.

22 Conceptual impulsivity and inductive reasoning

J. Kagan, L. Pearson and L. Welch

A series of investigations (Kagan, 1965a; 1965b; Kagan, Rosman, Day, Albert and Phillips, 1964) has demonstrated the temporal stability and intertask generality of a tendency toward fast or slow decision times to problems with high response uncertainty. Some children impulsively report the first hypothesis that occurs to them, and this response is often incorrect. The reflective child, on the other hand, delays a long time before reporting a solution and is usually correct. The reflective child considers the differential validity of alternative answers, makes fewer errors in reading prose or in recalling serially learned material, and persists longer with difficult tasks. The reflective child wants to avoid making an error and inhibits potentially incorrect hypotheses. The impulsive child seems minimally concerned about mistakes and makes his decisions quickly. Individual differences in time taken to evaluate a solution hypothesis touch the problem-solving process in two places: (*a*) selection of a solution hypothesis to act upon and (*b*) evaluation of the quality of the solution obtained. The chronology of a typical problem-solving sequence passes through four phases. In phase one, the problem is decoded (read or listened to) and comprehended. In phase two, a 'best' hypothesis is selected to act upon. Some evaluation is appropriate at this point. In phase three, the hypothesis is implemented, and, once again, there should be some evaluation of the validity of the solution arrived at. Finally, in phase four, the solution is reported to an external agent.

It is reasonable to predict that impulsive children will make more errors than reflective children of similar verbal ability on inductive-reasoning problems, for they should be prone to select an answer that was not carefully evaluated.

This paper summarizes an empirical test of this prediction. A subordinate aim of the paper was to explore further the temporal characteristics of eye scanning in visual-discrimination tasks. This is a relevant issue in this context because the most sensitive index of the reflection-impulsivity dimension is a test called Matching Familiar Figures (MFF). In this test, the child is shown a picture of a familiar object, called the standard, and six similar stimuli, only one of which is identical to the standard. Figure 1 illustrates two sample items from the test.

The child is asked to select the one stimulus that is identical to the standard, and the standard and six variations are always available. Previous work indicated that the children tend to orient their head and eyes toward

---

Source: *Child Development*, 1966, *37*, 583–94.

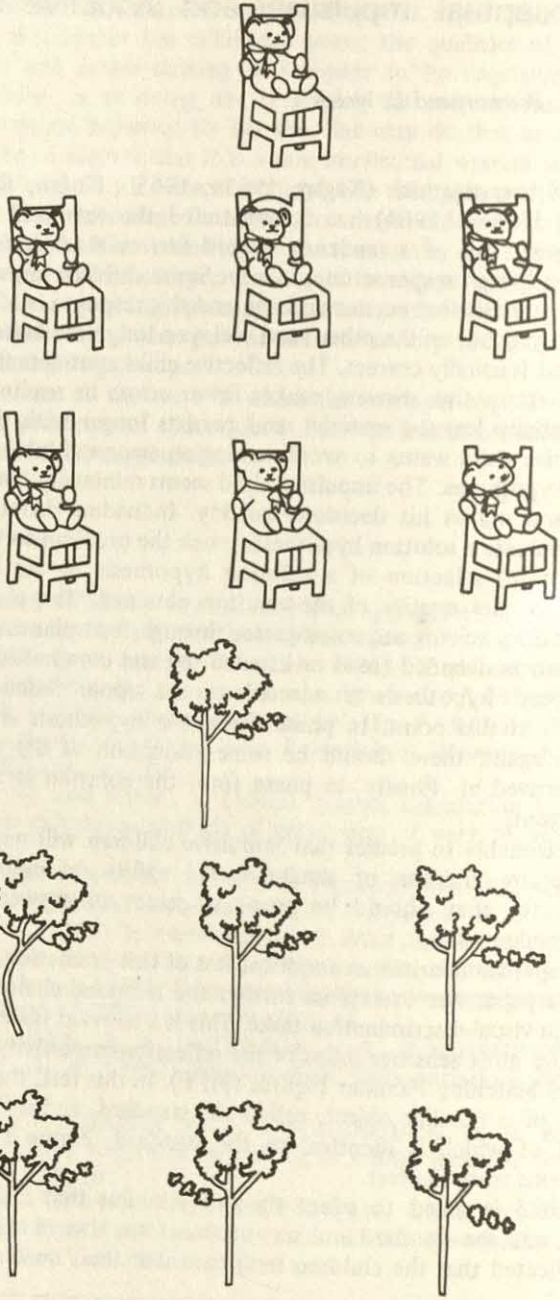


Fig. 1 Sample item from Matching Familiar Figures Test

the standard about every three seconds. This value of three seconds tends to be invariant across various response latencies, sex, and age of child. That is, a response time of 3 seconds usually is accompanied by one orientation of the standard, a response time of eleven seconds by three fixations; a response time of thirty-five seconds by ten fixations of the standard. Earlier studies did not permit a fine-grained analysis of the temporal characteristics of the scanning process. In order to obtain further data on this issue, assessment of head-eye orientations of the standard was made a part of this investigation.

Method

Subjects. Seventy-nine boys and 76 girls in the first grade in Newton, Massachusetts, were seen individually by one of three female *E* [experimenters] for two test sessions in the fall of the year. The test batteries consisted of measures of reflection-impulsivity, assessment of performance on inductive-reasoning problems, and the information and vocabulary scales from the Wechsler Intelligence Scale for Children (WISC). The first session included the following four tests:

Matching Familiar Figures (Figure 1).–MFF was administered according to a standard procedure. The child was asked to select the one variant that was identical to the standard. Errors and response time were recorded for each of the twelve test items. In addition, the *O*s [observers] recorded each time the *S*'s [subject's] head and eyes oriented to the standard and the duration of the orientation. All variables were recorded on an event-recorder. The major measures derived were (*a*) number of fixations prior to the *S*'s first offering of a solution hypothesis, (*b*) duration between each successive orientation of the standard prior to the first hypothesis, (*c*) number of orientations between *S*'s first and second solution hypotheses. The other three variables scored were total number of errors on the test, response latency to the first hypothesis (hereafter called MFF response time) and response latency between the offering of the first and second hypotheses. Thus six variables were derived from the MFF.

Haptic Visual Matching.–This task was administered in order to affirm the generality of a reflective or impulsive tempo for first-grade children. Earlier work demonstrated this generality for older subjects. In the Haptic Visual Matching Test (HVM) the child first explored haptically a three-dimensional wooden form to which he had no visual access. After his exploration, he selected from a visual array the one visual pattern that represented the form he had explored haptically. The task contained twenty stimuli, ten geometric forms, and ten familiar objects. The child was told to palpate the wooden form as long as he wished and to withdraw his hand from behind a curtain when he was finished. He was then shown the card illustrating five ink-line drawings and asked to select the visual pattern corresponding to the form he had palpated. The major variables scored were exploration time of the standard, response latency to first selection from the visual array, and number of errors.

Picture Completion Reasoning Test.–The child was shown a page containing three pictures that told the beginning of a sequence or story. A contiguous page contained four pictures, one of which was the most logical completion of the three-sequence story. The child was asked to select the one picture from the four that was the 'next thing that happened in the story'. Figure 2

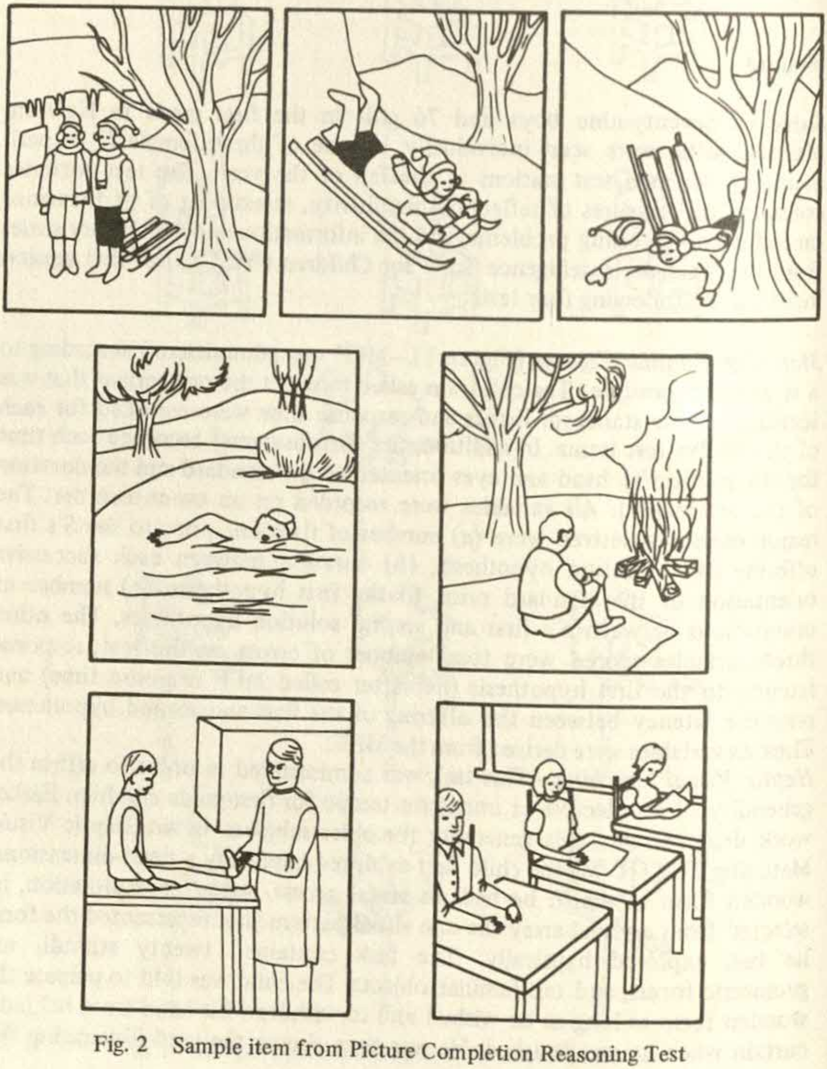


Fig. 2 Sample item from Picture Completion Reasoning Test

illustrates a sample item from this test. The *E* recorded response time to first selection and whether the response was correct. There were two practice and twelve test items administered.

Vocabulary level.–The vocabulary scale of the WISC was administered according to standard instructions.

The second test session also included four procedures.

Extrapolation Reasoning Test.–In this second reasoning test *S* was shown a series of stimuli that were linked by a principle. The child was also given a set of alternative stimuli and had to select the one that would continue the series according to the principle illustrated. Figure 3 illustrates a sample item. The directions to the child were as follows: 'You see this line of things here. One of the things down here should come at the end of this line. Now what do you think comes next? Point to the one that comes next'. The *E* recorded response time to the child's first selection and whether the response was correct.

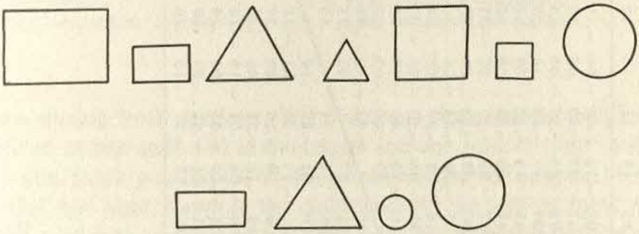


Fig. 3 Sample item from the Extrapolation Reasoning Test

Guessing objects.–In the final test of inductive reasoning the child was told three characteristics or attributes of an object and asked to guess the object. The instructions were as follows: Now we will play a guessing game. I am thinking of something and you will have to find out what I am thinking about. I will tell you a little about the thing I am thinking of so you will have some idea of what it might be. Let's do one for practice. I am thinking of something that has four feet, a furry tail, and barks. What is it?'

After two practice items the child was given twelve test items. Sample items included: 'What has wheels, a horn and doors?' 'What is slippery, smells nice, and you use it to clean?' 'What is orange, hot, and makes smoke?' 'What is round, small, and you find it on a shirt?'

The *E* recorded the response time to the child's answer and whether it was correct.

Information scale.–The information subscale from the WISC was administered to each child.

Self-evaluation.–The child was asked to evaluate his performance on the

Table 1 *Intercorrelations among major variables**

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	
1		−.56	−.55	−.27	−.57	−.58	.22	−.27	−.45	−.22	−.02	.25	.09	−.16	.46	−.04	−.15	.27	−.30	−.19	.27	−.21	−.13	−.29	1
2	−.63		.81	.56	.66	.47	−.40	.44	.51	.33	.05	−.25	−.03	.37	−.40	.00	.28	−.35	.40	.39	−.08	.18	.16	.08	2
3	−.67	.96		.19	.57	.62	−.42	.24	.51	.35	.11	−.17	.03	.23	−.39	−.05	.25	−.33	.24	.24	−.11	.12	.19	.10	3
4	−.45	.75	.65		.29	−.04	−.27	.35	.25	.04	.00	−.21	−.04	.17	−.22	.14	.25	−.19	.32	.36	−.14	.10	.02	.06	4
5	−.48	.61	.63	.52		.71	−.30	.18	.38	.22	−.02	−.14	−.01	.13	−.30	.00	.18	−.24	.54	.50	−.06	.04	.02	.03	5
6	−.41	.50	.56	.36	.92		−.24	.13	.35	.19	−.02	−.18	−.11	.00	−.30	.08	.17	−.09	.19	.20	−.14	−.01	.16	.16	6
7	.49	−.46	−.51	−.26	−.23	−.26		−.15	−.40	−.20	−.07	.29	−.17	−.09	.39	.05	−.21	.41	−.11	−.15	.07	−.18	−.27	−.13	7
8	−.30	.51	.42	.47	.26	.14	−.33		.39	.15	.09	−.24	−.06	.09	−.33	.01	−.05	−.16	.19	.29	.05	.06	.07	.03	8
9	−.49	.76	.67	.57	.23	.16	−.53	.70		.04	.10	−.33	.01	−.02	−.44	−.06	.15	−.33	.16	.31	.07	.18	.14	.05	9
10	−.33	.55	.44	.48	.09	.05	−.37	.49	.77		.28	−.20	.16	.25	−.29	−.08	.04	−.29	.15	.20	−.08	.11	.16	.20	10
11	−.34	.27	.26	.36	.18	.12	−.24	.24	.36	.50		.00	.32	.06	−.10	−.01	−.07	−.09	−.02	.16	.06	.03	−.03	.02	11
12	.37	−.45	−.47	−.37	−.28	−.21	.51	−.34	−.47	−.37	−.42		.10	−.10	.45	−.17	−.21	.31	−.09	−.14	.14	−.46	−.32	−.37	12
13	−.08	−.12	−.09	−.12	.16	.18	−.06	.03	−.05	.05	.22	−.09		.12	.08	.07	.17	−.08	.04	.05	.12	−.21	−.14	−.09	13
14	.02	.12	.14	.12	.20	.25	.00	.15	.03	.16	.20	−.03	.47		−.33	−.19	.11	−.21	.13	.25	.11	.10	.13	.11	14
15	.38	−.29	−.33	−.25	−.10	−.14	.45	−.31	−.36	−.31	−.30	.38	−.01	−.27		.03	−.21	.48	−.21	−.27	.00	−.44	−.43	−.35	15
16	−.15	.00	.08	.01	−.06	−.01	−.11	.00	−.02	−.12	.05	−.05	−.08	.06	−.21		.53	.19	.04	−.07	.01	.13	−.03	.14	16
17	−.38	.54	.53	.36	.10	.02	−.45	.39	.61	.33	.18	−.27	−.10	.01	−.27	.10		−.21	.17	.14	−.04	.07	−.07	.03	17
18	.15	−.29	−.28	−.29	−.18	−.12	.36	−.30	−.28	−.29	−.27	.37	.01	−.18	.35	.01	−.26		−.21	−.18	−.06	−.27	−.09	−.01	18
19	−.22	.48	.43	.38	.19	.12	−.24	.30	.45	.42	.31	−.28	.00	.04	−.20	.04	.37	−.16		.68	.12	.01	−.08	.00	19
20	−.15	.28	.24	.18	.08	.07	−.22	.25	.36	.41	.30	−.21	.15	.16	−.23	.06	.24	−.20	.75		.22	.01	−.04	.01	20
21	.02	−.07	−.07	−.05	.12	.13	.03	−.01	−.09	−.14	−.09	.09	.04	−.06	.05	−.16	−.10	.08	−.15	−.09		−.09	−.13	−.06	21
22	−.22	.10	.12	.20	−.11	−.11	−.26	.08	.20	.04	.09	−.27	−.21	−.24	−.43	.01	.24	−.17	.01	.00	−.04		.41	.38	22
23	−.16	.17	.19	.15	.05	.05	−.23	.14	.22	.16	.15	−.18	.01	.05	−.42	.09	.16	−.08	.12	.22	−.06	.28		.68	23
24	−.20	.25	.28	.22	.07	.06	−.21	.12	.20	.15	−.05	−.19	−.16	.05	−.38	.13	.26	−.19	.09	.10	−.19	.33	.69		24

*Boys to right and above diagonal; girls to left and below diagonal; $r = .22$ for $p < .05$, two-tailed test; $r = .29$ for $p < .01$, two-tailed test.
For Key, see opposite page.

Key to Table 1

(1) MFF: errors;
(2) MFF: response time to first hypothesis;
(3) MFF: fixations prior to first hypothesis;
(4) MFF: duration between fixations;
(5) MFF: response time between first and second hypothesis;
(6) MFF: fixations between first and second hypothesis;
(7) HVM: errors;
(8) HVM: exploration time
(9) HVM: response time;
(10) Picture Completion: response time to incorrect;
(11) Picture Completion: response time to correct;
(12) Picture Completion: errors;
(13) Extrapolation: response time to incorrect;
(14) Extrapolation: response time to correct;
(15) Extrapolation: errors;
(16) Guessing: response time to incorrect;
(17) Guessing: response time to correct;
(18) Guessing errors;
(19) Self-evaluation: response time about performance on test;
(20) Self-evaluation: response time, all attributes;
(21) Self-evaluation: actual rating, performance;
(22) WISC verbal IQ;
(23) Father's education;
(24) Mother's education;

tests given during both sessions. The child was shown a ten-inch line with demarcations at the ends and in the center and was told: 'I want to find out how well you think you did on all the games we played today and last time. Look at this line here. Suppose the child who did the best on these games is up here [*E* points to one end of the line] and the child who did the worst is here [*E* points to the other end of the line] and the one who did medium is here. Where would you be? Here is the pencil; make a mark to show how well you did'.

The *E* recorded the latency to the child's response. This procedure was repeated for three other attributes (the child who thinks fastest versus slowest; the child who is most afraid of making a mistake versus the child who is least afraid; the boy who is strongest or the girl who is prettiest). Response time to each of these answers was recorded.

A set of twenty-four variables was derived from the two assessment sessions, and these were subjected to an intercorrelational analysis.

Results and discussion

Table 1 presents the product-moment correlations among the twenty-four major variables. Several major conclusions emerge from this matrix of coefficients. First, there was good generality for both response time and errors across MFF, HVM, and some of the reasoning tests, with greater consistency among girls than boys. The coefficients were generally higher for girls than for boys when response time on MFF was correlated with the other response-time measures. Moreover, the usual negative correlation between response time and errors on MFF or HVM was consistently larger in magnitude for girls than for boys. These sex differences match earlier findings and affirm the conclusion that girls display greater intertask consistency for the reflective-impulsive attitude.

It is important to note that response time to *incorrect* responses on the Picture Completion Reasoning Test was the best correlate of MFF response time, whereas on the extrapolation Reasoning and Guessing Objects Tests response time to correct responses displayed the highest association with MFF response time. This particular pattern of correlations was not anticipated. A post hoc study of the items suggests that the picture-completion questions were easier than those on the extrapolation or guessing tests. Some items on Picture Completion Reasoning Test were so easy that the child knew the answer immediately and response uncertainty was minimal. Response uncertainty was present only for the difficult items, and on these items the child's conceptual tempo was revealed. On the Guessing Objects Test, the child had to generate his own alternatives, and more of the items were difficult. It is likely that long response times on this test reflected inability to produce any solution hypothesis rather than a labored consideration of alternatives. As a result, response latency to correct solutions should be a better correlate of MFF response time than latency to incorrect solutions. The Extrapolation Reasoning Test also contained many difficult problems. There was sufficient uncertainty in each item so that response time to correct answers may have been the best index of the reflection-impulsivity dimension. Long response times probably reflected confusion regarding the correct answer or how to obtain it. Informal questioning of the children after the test indicated that many children found the extrapolation test difficult.

As expected, response times to MFF and HVM were the most highly correlated ($r = 0.51$ [boys] and 0.76 [girls]), for these tasks are similar in their conceptual requirements. As in earlier work, errors on MFF or HVM were negatively correlated with WISC verbal-scale scores (correlations around 0.20 to 0.25), but response time to MFF or HVM was independent of verbal ability (r's ranged from 0.08 to 0.18).

Decision time for self-evaluation was longer for reflectives than impulsives. Children with long response times on MFF took longer to decide about the quality of their performance on the tests as well as the other personal attributes. This result has implications for investigators who use self-rating information in personality research. The time taken to consider a self-rating co-varies with its accuracy, and since reflective children take longer to decide about self-evaluative statements, their self-descriptions may be more accurate.

The results summarized above treated MFF response time as a single, continuous variable and indicated that *S*s who had long response times on MFF (i.e. reflective *S*s) made fewer errors on the inductive-reasoning tests. Although there is a strong negative correlation between response time and errors on MFF, there is a small group of *S*s who show long response times and many errors or short response times and few errors. Thus, an alternative operational definition of the reflection-impulsivity dimension combines response time and errors. These purer groups were created by classifying a *S* as impulsive if he were both above the median on MFF errors and below the median on MFF response time for his sex. A reflective child was below the median on MFF errors and above the median on MFF response time. This categorization yielded 25 impulsive and 30 reflective boys, 26 impulsive and 26 reflective girls. Point biserial correlations were computed between the dichotomized reflection-impulsivity variable and the remaining continuous variables.

The magnitude of the relations between this index of reflection and the remaining variables was not markedly different from the correlations using response time alone. The point biserial correlations with errors and response time on HVM were – 0.48 and + 0.45 for boys; – 0.43 and + 0.45 for girls. The relations with the picture-completion, extrapolation, and guessing tests were approximately of the same magnitude as those reported in Table 1 for the whole group. The most consistent set of relations involved error scores. Reflective children (i.e. long response times and few errors on MFF) displayed fewer errors on all three inductive reasoning tests (five of the six correlations were statistically significant). There was also a slight positive correlation (correlations around 0.25 to 0.30) between a reflective attitude and the educational level of each parent.

It is interesting to note that the reflective boys, who performed better on the objective tests, evaluated themselves as slightly more competent than their impulsive counterparts. However, these reflective boys also delayed longer before offering the evaluation of their performance. Reflective boys gave a more thoughtful judgment of their ability, and this may be one reason that they were more accurate. There was no relation among girls between a reflective attitude and self-rated ability, although, in fact, the reflective girls performed better than the impulsives.

In sum, the data support the generality of a reflective attitude and verify the prediction that reflective children would make fewer errors on inductive-reasoning tests. When partial correlations were computed between MFF response time and errors on picture-completion, extrapolation, and guessing tests, with WISC verbal ability partialed out, the resulting correlations were – 0.20, – 0.36, and – 0.32 for boys and – 0.44, – 0.28, and – 0.28 for girls (all coefficients larger than 0.23 are significant at $p <$ 0.05 for two tails). These data indicate that the relation between reflection and accurate performance on inductive-reasoning tests holds even when verbal ability is statistically controlled. The categorization of a child as reflective or impulsive implies a combination of speed and accuracy of response. Thus, a psychologically more reasonable index of reflection-impulsivity should combine response time and errors as we did in this second analysis. Whenever one has a large sample this is the preferred procedure. However, the negative correlation between response time and errors on MFF is always strong, and we have never found qualitatively different results when we used response time alone or combined response time with errors as the index of the reflection-impulsivity dimension. When one has small samples or wishes to limit administration time by not requesting the child to offer additional guesses after his first error, it is not dangerous to use response time alone as an approximate index of this response disposition.

Eye fixations of the standard

The data on head-eye fixations support earlier findings indicating an unusually high relation between the number of fixations and response time prior to the first solution hypothesis. In earlier studies the correlations were between 0.90 and 0.95. In this study, the coefficients were 0.81 for boys and 0.96 for girls. In addition, the other variables behaved in a manner consistent with theoretical expectations. Children who made many fixations prior to the first hypothesis also made more fixations between their first and second hypotheses ($r = 0.62$ and 0.56). Moreover, response time prior to the first solution hypothesis was highly correlated with the latency between solution hypotheses 1 and 2, and each latency was highly associated with number of fixations.

Finally, the length of time the child studied the six variants (i.e. time between successive fixations and the standard) was highly correlated with the total number of fixations prior to the first hypothesis and with a reflective attitude. Reflective children not only made more fixations of the standard but they also spent more time studying the six variants. Impulsive children made fewer orientations of the standard and spent less time studying the variants. All phases of stimulus processing were faster for the impulsive children.

It is of interest to note that the distributions of time delays between successive fixations of the standard were leptokurtic, with a node at two seconds and the vast majority of durations falling between two and three seconds. That is, independent of number of fixations or response time prior to the first hypothesis, the children typically tended to initiate a fixation of the standard every 2–3 seconds. Figure 4 illustrates the distribution of these time delays for the impulsive and reflective children (sexes pooled; using response time and errors as index of reflection-impulsivity). The peak value is the same for both reflective and impulsive subjects, between two and three seconds. However, occurrence of unusually long delays between successive fixations was more frequent for reflective subjects, whereas unusually short time delays were more frequent for impulsives.

The favored interpretation of these data is that orientation of the standard may be linked to the respiration of the child. Children this age usually have a respiratory rate between two and three seconds, and it is possible that the act of inspiration triggers a tendency to lift one's head and orient to the standard. Experiments in progress are directed at assessing the validity of this interpretation.

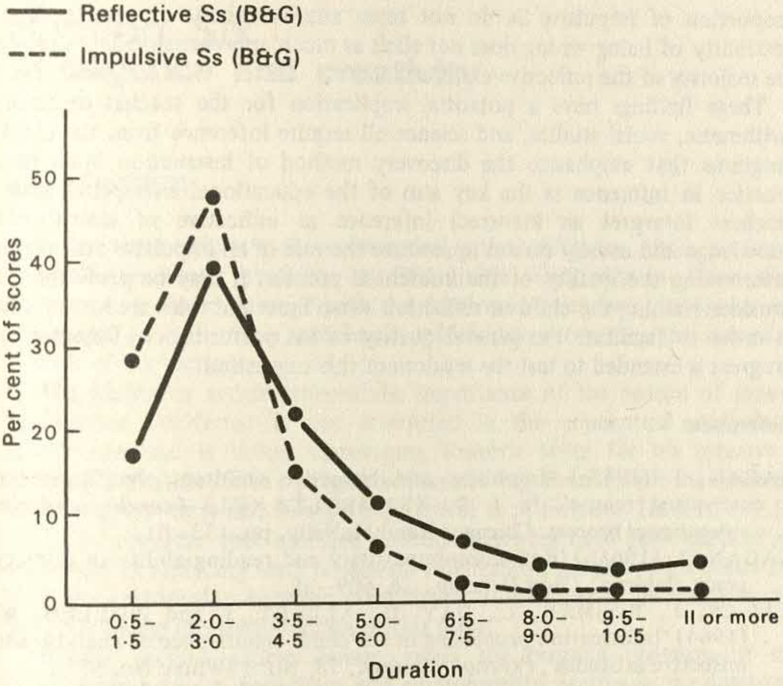


Fig. 4 Temporal durations between successive fixations of the standard

General discussion

The data support the assumption of a general tendency toward reflective or impulsive responding in problems with high response uncertainty for children as young as six years of age. The findings demonstrate further that a reflective tendency generalizes to inductive-reasoning tasks that contain response uncertainty. Impulsive children make more errors in inductive-reasoning problems because they do not pause to evaluate the quality of their inferences. The impulsive child responds quickly in situations where inferences are required; he seems to report the first reasonable idea that occurs to him. His failure to censor or to evaluate critically the quality of his inferences may be due to two factors. On the one hand, the impulsive child may have difficulty placing effective inhibitions on tendencies toward action. He may find it hard to block the urge to blurt out an answer even though he is not sure it is correct and even though he would like to be correct. On the other hand, informal observations indicate that a fair proportion of impulsive Ss do not seem anxious about a mistake. The possibility of being wrong does not elicit as much apprehension as it does in the majority of the reflective children.

These findings have a potential implication for the teacher or tutor. Arithmetic, social studies, and science all require inference from the child. Programs that emphasize the discovery method of instruction insist that practice in inference is the key aim of the educational enterprise. Many teachers interpret an incorrect inference as indicative of insufficient knowledge and usually do not appreciate the role of an impulsive attitude in determining the quality of the inferential process. It may be profitable to consider training the child in reflection when facts and rules are introduced in order to facilitate the general quality of his performance. Research in progress is intended to test the wisdom of this suggestion.

References

KAGAN, J. (1965a) 'Impulsive and reflective children: significance of conceptual tempo'. In J. D. KRUMBOLTZ (Ed.) *Learning and the educational process.* Chicago: Rand McNally, pp. 133–61.

KAGAN, J. (1965b) 'Reflection-impulsivity and reading ability in primary grade children', *Child Develop., 36,* 609–28.

KAGAN, J., ROSMAN, L., DAY, D., ALBERT, J. and PHILLIPS, W. (1964) 'Information processing in the child: significance of analytic and reflective attitudes', *Psychol. Monogr., 78,* No: 1 (Whole No. 578).

Section Six
Intelligence and creativity

Introduction

The first three articles in this section deal with intelligence, while the fourth and fifth are about creativity; the fifth article, 'Creativity and Curriculum Structure', is also concerned with the relationship between intelligence and creativity, so it forms a conclusion which draws together the two main threads of the section.

The McNemar article stresses the importance of the notion of general intelligence. McNemar is not interested in the conceptual analysis of intelligence and is rather disparaging towards Miles for his interest in defining it. Rather, he is a pragmatist, concerning himself with the structure of intelligence tests and how socially useful such tests are. He does not, for example, consider that multiple aptitude batteries have been particularly successful in replacing tests of general intelligence. As part of his concern is with the relationship between creativity and intelligence, his article is also linked with the creativity section of this topic.

In the second article Jensen raises the burning question of the relationship between hereditary and environmental factors in the determining of intelligence. It is often difficult, many would say impossible, to separate educational from political and ideological issues, and this is particularly true of the heredity–environment debate.

Jensen looks at the issue in terms of black and white people's intelligence, considering such questions as the effect of malnutrition on intelligence and the nature of verbal deprivation. In this article he is not so much interested in presenting the hard data behind his theories (this he has done elsewhere), but with discussing some of the fundamental ideas as objectively as possible.

In the third article Stinchcombe continues the heredity–environment controversy. Whereas Jensen stresses the importance of genetic factors, Stinchcombe emphasizes the importance of environmental factors. Indeed, much of Stinchcombe's article is concerned with refuting Jensen's interpretation of the evidence (although the Jensen article he is most concerned with is an earlier one, not included in this Reader). Stinchcombe stresses how little is known of the complexity of the environment and argues that, for this reason, statements about the proportional effects of heredity and environment are, at the current stage of knowledge, of little value. Much of Stinchcombe's article is concerned with the nature of the environment, and particularly with propagating a cumulative (and correspondingly complex) view of it, as opposed to a simplistic view.

The article by Jackson and Messick attempts a conceptual analysis of the term 'creativity' and tries to relate various aspects of this concept to such things as the differing personality characteristics of the learner. The compartments into which this article divides the various components are too comfortable, but it does attempt a useful analysis of the ways in which creativity can be defined, other than by using an operational definition.

The final article in this section, by Ogilvie, deals with the effects of school environment on creativity, and also with the complexity of the interrelationships between intelligence, creativity and curriculum structure. The article, which falls into two main parts, reflects Ogilvie's concern to connect sound, empirical techniques to a firm theoretical background.

The first part is a review of the literature dealing, in particular, with the relationship between 'formal' and 'informal' schools and creativity. The second part describes a piece of research on creativity carried out by Ogilvie himself. He bases his work on Mednick's Associationist Theory of Creativity and presents the results obtained when he adminstered five creativity tasks (based on the work of Guilford and that of Wallach and Kogan) to ten- and eleven-year-olds in five schools with different environments.

Victor Lee

23 Lost: our intelligence? Why?

Q. McNemar

The Greeks had a word for it, but the Romans had a word with better survival properties. Regardless of the word, what is now called intelligence has been talked about for at least 2000 years. And as long as 2000 years before the advent of attempts to measure intelligence, there seems to have been recognition of the fact that individuals differ in intellectual ability.

The earlier attempts at measuring were based on either of two quite distinct conceptions: the Galton–Cattell idea that intellectual ability manifests itself in simple, discrimination functioning, and the Binet notion that cognitive ability reflects itself in more complex functioning. The Binet concept proved to be more fruitful, and by 1925 there was on the market, in addition to various versions of the Binet scale, a flood of group tests of so-called general intelligence.

A few words about definition may be in order. First, it might be claimed that no definition is required because all intelligent people know what intelligence is–it is the thing that the other guy lacks. Second, the fact that tests of general intelligence based on differing definitions tend to intercorrelate about as highly as their respective reliabilities permit indicates that, despite the diversity of definitions, the same function or process is being measured–definitions can be more confusing than enlightening. Third, that confusion might have been anticipated is evident from recent reexamination of the problem of definition by Miles (1957). This British chappie found himself struggling with the awful fact that the word 'definition' itself has twelve definitions. Perhaps the resolution of this problem should be assigned to the newly formed Division of Philosophical Psychology, or maybe the problem should be forgotten since psychologists seem to have lost the concept of general intelligence.

Why has the concept been abandoned? Was it replaced by something else? By something better? Must we admit that the millions who have been tested on general intelligence tests were measured for a non-existent function? If it is possible that the notion of general intelligence is not lost but merely gone astray, in what corners of what psychological fields should we search for it?

Source: *American Psychologist*, 1964, *19*, 871–82. Abridged version reprinted from BUTCHER, H. J. and LOMAX, D. E. (1972) *Readings in Human Intelligence*. London: Methuen Educational, 40–56.

Reasons for discarding the idea of general intelligence

Apparently one reason why concepts are either discarded or modified beyond recognition is that too much is claimed for them. Among the supposed strikes against general intelligence are the following: the earlier false claims about IQ constancy; prediction failures in individual cases; unfounded claims that something innate was being measured by the tests; equally unfounded assertions that nothing but cultural effects were involved; the bugaboo that IQ tests reflect middle-class values; the notion that an IQ standing fosters undesirable expectations regarding school achievement; the idea that IQ differences are in-compatible with democracy and lead to educational determinism; and, finally, the great stress on general intelligence caused us to ignore other possible abilities.

This last point leads us right into the problem of factor analysis. Spearman died in battle defending his theory of *g*. Under pressure he reluctantly conceded that factors other than *g* might exist, and he frequently said, in effect, I told you so as long ago as 1906. Actually, Spearman was on the run before the invention of modern factor analysis, but it was not until Thurstone's (1938) first major application of his centroid factor method that Spearman's *g* became, seemingly, non-existent. Thurstone said, 'we have not found the general factor of Spearman' and 'We cannot report any general common factor in the battery of fifty-six tests'. As anticipated by some, Spearman was not prone to admit defeat. He reworked Thurstone's data and a *g* was found, plus some group factors. He charged that Thurstone's rotational process had submerged the general factor.

American factorists found Thurstone convincing. The description of abilities in terms of seven primaries was an attractive package. The so-called primaries were more amenable to specific definition than the old hodge-podge called general intelligence. Despite the fact that Thurstone was able to replicate his findings on samples from two other populations, thus giving credence to his method and results, there were a couple of events that led to some turbulence in his seven-dimension rarefied atmosphere. The first of these was a minor study, by one of his own students, based on the intercorrelations of 1916 Stanford–Binet items, in which the *g* refused to be rotated out. But rather than admit that this might be some kind of general intelligence, the author renamed it 'maturational level'. Incidentally, this illustrates the first cardinal Principle of Psychological Progress: *Give new names to old things.*

The second disturber of the neat little set of primaries, sans a *g*, resulted when Thurstone took the next logical step, that of constructing tests to measure the primaries. It was found that the primaries were themselves intercorrelated whereas it had, at the time, been expected and hoped that they would be independent. The Thurstones (1941) readily admitted that a general factor was needed to explain the interrelatedness of the primaries.

This eventually led to the idea of oblique axes, which axes were regarded as representing the primaries as first-order factors, whereas the general factor pervading the primaries was dubbed a second-order factor. It began to look as though Spearman was being revisited, except for the little matter of labelling: anything called second-order could not possibly be regarded as of much importance. Furthermore, it could always be said that, in the ability domain, it is less difficult to attribute psychological meaningfulness to first-order than to second-order factors, so why pay much attention to the latter? Thus it was easy for most American factorists to drop the concept of general intelligence and to advocate that tests thereof, despite their proven usefulness over the years, should be replaced by tests of the primaries. Hence the emergence of differential aptitude batteries, about which more later.

Meanwhile, our British cousins did not tag along with the factor methods preferred on this side of the Atlantic. After all, it is possible to use factor methods that permit a sizeable general factor, if such exists, to emerge as the very first factor. Being first, it is, presto, the most important, as indeed it is as a factor explaining, for the starting battery as a whole, more variance for more tests than attributable to any American-style primary factor. The methods preferred by the British also yield group factors, apt to bear the same name as the primaries, but of attenuated importance. Apparently the British are sceptical of the multitude of ability factors being 'discovered' in America. The structure of intellect that required 120 factors may very well lead the British, and some of the rest of us, to regard our fractionization and fragmentation of ability, into more and more factors of less and less importance, as indicative of scatterbrainedness. This statement presumes that intellectual abilities are brain-centred.

In practically all areas of psychological research the demonstration of trivially small minutiae is doomed to failure because of random errors. Not so if your technique is factor analysis, despite its being based on the correlation coefficient—that slipperiest of all statistical measures. By some magic, hypotheses are tested without significance tests. This happy situation permits me to announce a Principle of Psychological Regress: *Use statistical techniques that lack inferential power.* This will not inhibit your power of subjective inference and consequently will progress you right back to the good old days when there was no strangling stat or sticky stix to make your insignificant data insignificant.

It may be a long time before we have an ivory tower, strictly scientific resolution of the issue as to whether a scheme involving primary abilities plus a de-emphasized *g* is preferable to one involving an emphasized *g* plus group factors. With bigger and better computers we will have bigger, though not necessarily better, factor analytic studies, but it seems unlikely that such further studies will, in and of themselves, settle the issue under discussion. Until such time as some genius resolves the broader question, so ably discussed by Cronbach (1957), of the place, if any, of correlational method in a science that aspires to be experimental, we may have to turn to the criterion of social usefulness as a basis for judging whether it is wise to discard general intelligence. Like it or not, much of our heritage in this area is that earlier workers, from Binet on, had as their motivation the solution of social problems, and currently many in the area have a similar motivation.

The bearing of social usefulness

In practice, if you believe that the concept of general intelligence has outlived its usefulness, you may choose from among several differential, or multiple, aptitude batteries, which will provide measures of some of the so-called primary mental abilities. If you happen to believe that there is something to general ability, you can find tests to use. The novice looking for these latter tests may have to alert himself to the first Principle of Psychological Progress–the test labels may have changed from 'general intelligence' to 'general classification' or 'scholastic aptitude'. If you enjoy riding the fence, you might become a devotee of the practice of the College Board, and others, and measure just two abilities: Verbal and Quantitative.

This is certainly not the place to review the voluminous literature that amply demonstrates the practical utility of tests of general intelligence. Nor is it the place to catalogue the misuses of the Stanford–Binet for purposes which Terman never claimed for it, or the misuses of the Wechsler scales for the purposes which Wechsler *has* claimed for his scales. Neither the Binet nor the Wechsler provides a factorially pure, unidimensional measure of a *g*. The current Stanford–Binet was in reality constructed too early to benefit from the implication of factor analysis for test purity, whereas the Wechsler scales were based on the impossible premise that ten or eleven subtests can simultaneously provide diagnostic subscores and a meaningful total score of the many group tests that appeared between 1920 and 1945 it can be said that few, if any, provide unidimensional measures of general intelligence. The chief difficulty is that most of them lead to a total score based on a mixture of verbal and mathematical material. Thus, with two main sources of variance, marked qualitative differences can exist for quantitatively similar total scores. The College Board–Educational Testing Service people have justifiably refrained from giving a total score involving verbal plus math, but there are those who question the usefulness of the Board's math score and there are those who criticize the Educational Testing Service for failing to change over to a differential aptitude battery. [. . .]

Just how successful have the multitest batteries been? Since by far the most extensive social use of tests has been, and continues to be, in the schools, let us look at the evidence of validity studies therein. As indicated previously, little is known about the predictive usefulness of some of the seven batteries discussed above. The DAT [Differential Aptitude Tests] of the Psychological Corporation is the only battery for which adequate predictive (and concurrent) validity data, derived from school sources, are available. It is also the battery that has fared best in the hands of the test reviewers; therefore if we allow the case for differential batteries to rest thereon, we will be looking at the best. So, what is the story?

Recall that the hoped-for advantage of a multitest battery over the old-fashioned general intelligence test was that it would have greater predictive power, a power which could manifest itself in higher validity coefficients for specific subject matter and, perhaps, for overall achievement. It was hoped that such a battery would be truly differential in that particular factors (or subtests) would correlate higher with achievement in some areas than in other areas. Presumably each factor (or subtest) should have unique usefulness. If a battery were truly differential, it would be a boon at school guidance personnel.

Now the manual of the DAT of the Psychological Corporation contains a staggering total of 4096 (yes I counted'em) validity coefficients. With such a large pool to draw from, one could by gracious selection 'show' that the DAT is the answer to the prayer of every counsellor, male or female, or by malicious selection one could 'prove' that the DAT is far worse than any test ever published. The validity coefficients range all the way down to–0.37, which is presumably a chance deviation downward from 0, and all the way up to 0.90, which is likely as not a chance deviation downward from unity. But ranges tell us nothing. After a careful perusal of the 4096 correlations, it seems safe to summarize DAT validities as follows:

1. Verbal Reasoning (analogies to most of you) is the best single predictor; Language Usage, as represented by a sentance test dealing with grammar and word usage, and admittedly more achievement than aptitude, is a close second.
2. Numerical Ability, as measured by a test of simple arithmetic operations, designed to tap arithmetic reasoning without the usual verbal component, is the best predictor of achievement in school mathematics. It does not, however, correlate as well with grades in science as does Verbal Reasoning.
3. Aside from the Numerical Ability test, the only other test that shows differential power as a predictor is the Spelling test–if you cannot spell you may have trouble learning shorthand.
4. The remaining five tests in the battery simply fail to show compelling evidence that they are good in the differential predictive sense. For the Mechanical Reasoning and the Clerical Speed and Accuracy tests this may be understandable in that little of school curricula for Grades 8 through 12 requires such abilities, but one would expect that Abstract Reasoning and Space Relations would fare better than they seem to.

Such data as we have been able to locate for the other six multitest batteries tend to support these findings on the DAT. Aside from tests of numerical ability having differential value for predicting school grades in math, it seems safe to conclude that the worth of the multitest batteries as differential predictors of achievement in school has not been demonstrated. Incidentally, the fact that the Verbal and Numerical tests stand out as the only two useful predictors tends to provide some support for the Educational Testing Service–College Board practice of providing scores for just these two abilities.

And now we come to a very disturbing aspect of the situation. Those who have constructed and marketed multiple aptitude batteries, and advocated that they be used instead of tests of general intelligence, seem never to have bothered to demonstrate whether or not multitest batteries provide better predictions than the old-fashioned scale of general intelligence. Be it noted that we are not discussing experimental editions of tests. Some may say that in so far as a test publisher provides validity data for a new battery it is not necessary to show that the validities are, for the given school condition, better than those of other tests. With this one can agree, but only in case no claims are made, explicitly or implicitly, regarding superior merits for the new battery.

It is far from clear that tests of general intelligence have been outmoded by the multitest batteries as the more useful predictors of school achievement. Indeed, one can use the vast accumulation of data on the validity of the Psychological Corporation's DAT to show that better predictions are possible via old-fashioned general intelligence tests. Consider the fact that a combination of the tests Verbal Reasoning (analogies) and Numerical Ability would be, in terms of content, very similar to many group tests of general intelligence. Consider also that an equally weighted combination of these two tests correlates in the mid–0.80s with the Otis S-A, Higher Form. Then, when you turn to a careful study of the empirical validities, as reported in the DAT manual, you will not be surprised at the outcome of the application of a little arithmetic, which leads to the definite conclusion that a simple unweighted combination of the Verbal Reasoning and Numerical Ability tests predicts as well as or, in most instances, better than any subtest taken singly, or in the differential sense.

The manual for the DAT contains the following statement (Bennett, Seashore and Wesman, 1952):

> Apparently the *Verbal Reasoning* and *Numerical Ability* tests can serve most purposes for which a general mental ability test is usually given in addition to providing differential clues useful to the counsellor. Hence, the use of the so-called intelligence test is apparently unnecessary where the *Differential Aptitude Tests* are already being used.

Anyone who disagrees with this quotation could, with better justification, say that an intelligence test can serve nearly all, if not all, the purposes for which a multiple aptitude battery is given in the schools because the former, in general, is a better predictor and because, as we saw earlier, the differential clues are too fragmentary to be of use to the counsellor. And there is a bonus: one classroom period of testing, compared to six periods. A second bonus: much less costly. A third bonus: fewer scores to confuse the already confused minds of most school counsellors.

Thus, we come to the conclusion that general intelligence has not been lost in the trend to test more and more abilities; it was merely misplaced by a misplaced emphasis on a hope that a lot of us, including the author, once entertained, a hope that in turn was based on a misplaced faith in factor analysis: *the* hope that factors, when and if measured, would find great usefulness in the affairs of society. By the criterion of social usefulness, the multiple aptitude batteries have been found wanting. Now, I have no desire to furnish ammunition for those test critics who would have us stop all testing merely because they find a trivially faulty item in a standardized test. At a time when there is shouting about the tyranny of the testers and the brass of the brain watchers, at a time when school people are showing resentment at the disruption caused by too many national testing programs, at a time when federal and state legislators are all too willing to write legislation that places restrictions on the use of tests, and at a time when both majorities and minorities are being denied the benefits of test-based guidance because certain well-intentioned persons fail to realize that scores for the underprivileged minorities are useful indices of *immediate*, or present, functioning–at a time when all these and other forces are operating to throw out the tests, it is high time for the profession to establish a bureau of standards to test the tests instead of coasting down a road that is tinged with some of the trappings of Madison Avenue. Better to have informed internal control than ignorant, hostile, external control.

Intelligence elsewhere?

Aside from the near loss of the idea that progress in school may depend on general intelligence, one wonders whether intelligence has come to be regarded as unimportant in other areas.

Any of you who have money invested in stocks and wish some reassurance regarding the intelligence level of business and industry managers should read Edwin Ghiselli's (1963) Bingham Lecture. His summary of his own work indicates that the average intelligence of those in the upper and middle management levels falls at the ninety-sixth percentile of the population. Thomas Harrell (1961) came to a similar conclusion. Furthermore, management level is correlated with intelligence–you can be too dumb to succeed as a manager. Also you can be too bright to be a managerial success! Now it must be admitted that little, if anything, is known about whether management success might be better predicted by measures of factor analytic defined abilities. On this you are free to guess–most of you will have already guessed my guess.

A one-by-one cataloguing of what we know or do not know about what abilities contribute to success within various occupational and professional groups would merely add to the dullness of this presentation, so let us turn to some of the more esoteric fields of psychology to see whether the concept of general intelligence has or has had any relevance. One such field, and a very broad one, is creativity. Anyone who peeks over the fence into this field is apt to be astonished at the visible chaos. The definition of creativity is confounded by the diversity of subareas within the field, the criterion problems are far from licked, and so little is known about the creative process that measuring instruments are, seemingly, chosen on a trial-and-error basis. [. . .]

Let us turn to a criterion-based study (Taylor, 1961). The criterion measures for creativity and productivity were based on the checking by supervisors of statements that had been scaled by Thurstone's equal-appearing interval method. Creativity and productivity, so gauged, correlated 0.69 with each other on a sample of 103 researchers (electronic scientists and engineers). For this same group, intelligence, as measured by the Terman Concept Mastery Test (CMT), correlated only 0.20 or less with the criteria. Two Psychological Corporation tests and an American Institute for Research test did a little better. Creativity is slightly more predictable than productivity. In so far as these two criteria are themselves valid, the findings indicate that within a group of research workers, precious little of the variance in creativity, and still less in productivity, can be predicted by the tests.

A study by MacKinnon (1962) based on criterion (rated) measures of performance was concerned with the creativity of architects. Although the author reports that *within* a creative sample the correlation is essentially zero between intelligence (CMT) and rated creativity, it is not clear from the context what is meant by 'within' sample. If this means within the sample of forty creative architects selected as the 'most creative' in the country, then we indeed have such a drastic restriction in range on the *criterion* variable that little, if any, correlation can be expected for any and all predictors. Now the author says, without presenting any evidence, that 'Over the whole range of intelligence and creativity there is, of course, a positive relationship between the two variables'. One wonders just what is meant by creativity in architectural journals. If judged creativity reflects engineering-structural innovation, then intelligence would likely be a correlate; if judged creativity depends on new artistic designs, then the intelligence component would likely be of less importance. It would seem that when the author says we 'may have overestimated. . .the role of intelligence in creative achievement', he should have included some marked qualifications as to what type of creativity he had in mind.

That such qualification is indeed necessary is supplied by a finding of still another investigator (Barron, 1963). For a group of highly creative writers it was estimated, by way of the Terman CMT, that their average IQ is about 140, which we interpret as meaning that a high IQ is a necessary, though not sufficient, condition for outstanding success as a writer. On the basis of his own studies and those of other persons, this same investigator suggests that 'over the total range of intelligence and creativity a low positive correlation' of 0.40 probably obtains. This sweeping generalization is for all areas of creativity.

And speaking of sweeping generalizations, consider the suggestion in a 1961 study (Holland, 1961) that 'we need to use non-intellectual criteria in the selection of students for scholarships and fellowships. The author did not say so, but presumably he meant in addition to intellectual ability; maybe he did not, since he had previously concluded that 'intelligence has little or no relationship to creative performance in arts and science . . . ' at the high school level. His data back up this conclusion, as might have been expected when correlations are based on groups restricted in range to the top 1 per cent!

If the foregoing examples of criterion-based studies of creativity seem to indicate that general intelligence is relatively unimportant for creativity, it should be remembered that drastic but unknown or unspecified curtailment of range exists for both ability and criteria. Why do correlational studies under such adverse circumstances?

Next we turn to a few studies of creativity which cannot be criticized because of restriction of range on the criteria–these studies simply avoid this problem by never having actual criterion information. The approach is to claim that certain tests, which typically are scored for novel responses or novel solutions to problems, *are* measures of creativity, with no evidence whatsoever that the tests have predictive validity for non-test, real-life creative performance. This bit of ignorance does not prove to be a handicap to those who think that creativity can be studied without the nuisance of obtaining criterion measures. We reluctantly accept the test-based criteria solely for the sake of seeing what happens to general intelligence as a part of the picture. Space permits only three examples.

We first note that general intelligence has not manifested itself as a correlate of so-called creativity tests in the factor analytic studies of creativity. The explanation of this is easily found–no measures of general intelligence are used in these studies. When discussing his plans for studying creativity, a certain author (Guilford, 1950) said that 'we must look well beyond the boundaries of the IQ if we are to fathom the domain of creativity. He went on to say, the conception 'that creative talent is to be accounted for in terms of high intelligence or IQ . . . is not only inadequate but has been largely responsible for lack of progress in the understanding of creative people. With a part of this one can agree, but does it follow that one should prejudge the role of general intelligence as a source of variance in creativity tests or factors derived therefrom? Does the failure to include an IQ test help one learn the extent to which one must go beyond the boundaries of the IQ to fathom creativity? Apparently the author, although willing to predict that the correlations between IQ and the many types of creativity tests 'are only moderate or low', was unwilling to include an IQ test for the sake of finding out. However, negation by omission is not very convincing.

That at least one test bearing the label 'creativity' is correlated more than moderately with IQ is evidenced by the value of 0.67 (average for boys and girls) for the carefully chosen sample of 15-year-olds in Project Talent (Shaycroft, Dailey, Orr, Neyman and Sherman, 1963). This sample-stable *r* (based on a total *N* of 7648) becomes 0.80 when corrected for attenuation.

In recent extensive study (Getzels and Jackson, 1962), already extensively criticized, creativity is defined as the sum of scores on five tests (median intercorrelation of only 0.28). Although the investigators use the sum score for most of their analyses, they do not bother to report the correlation of creativity, so defined, with IQ. From the published report I have ascertained (via the correlation-of-sums formula) that creativity and IQ correlate to the extent of 0.40 for the total of 533 cases. Now this *r* of 0.40 has been greatly attenuated because of three things: first, the usual measurement errors; second, the cases were highly selected in IQ (mean of 132); third, the IQs are a mixture from the Stanford–Binet, Henmon–Nelson and Wechsler Intelligence Scale for Children (the use of regression-estimated Binet IQs from the other two scales aggravates rather than improves the mixture). We deduce that intelligence and the creativity tests used here have far more common variance than the authors believe.

Much is made of the finding that the creativity tests tended to correlate higher than did IQ with verbal-content school achievements. Again the IQ comes in for an unfair drubbing because of the same mixture of IQ scores and, what is more pertinent, because of explicit selective curtailment on the IQ variable and only incidental selection on the creativity variable.

Of more importance to the present paper is the analysis, by these same authors, based on a high IQ group and a high creative group, these groups being selected as the top 20 per cent for each variable but excluding those who were in the top 20 per cent on both variables. These two selected groups were then contrasted on total school achievement (and a host of other variables that are of no interest here). The mean IQ for the high IQ group was 150 whereas the mean IQ for the high creative group was 127, yet the achievement means of the two groups were 'unexpectedly' equally superior to the school population mean despite the 23-point difference in mean IQ. The authors say that it '*is* quite surprising' that the high creativity group achieved so well. From this it is concluded that the 'creative instruments account for a significant portion of the variance in school achievement', and the subsequent argument implies that creativity is more important for ordinary school achievement than is the IQ. Now anyone who is at all familiar with a three-variate problem will not be 'unexpectedly' surprised at the foregoing results–indeed, if the authors had bothered to give the three basic correlations among the three variables (IQ, creativity and total school achievement) for the entire group, any person versed in simple multivariate analysis could deduce the results. Furthermore, he could deduce a further result (and this one has been overlooked by the critics) which might be unpleasantly surprising to the thesis of these authors: namely, the high IQ and the high creative groups did equally well in school achievement despite an unreported difference in mean creativity and that is of the same order as the much stressed difference in IQ. Utilizing the half-blind logic of the authors, one can say that creative ability is not as important as IQ for school achievement–just the opposite of their position.

Now the fact that seven of nine replications of this study confirm the original findings merely indicates that repetition of the same faulty design and false logic will lead to the same false conclusions. The design being used is such that, if two variables are equally correlated with a third, the conclusion will be reached that the two are actually unequally correlated with the third. This is the neatest trick of the decade for supplying educationists with an antidote for the IQ virus. I cannot refrain from saying at this point that, although discouraged, I am still hopeful that people who do statistical studies will first learn a modicum of elementary statistics!

Space does not permit a discussion of other studies in which creativity is defined in terms of test performance instead of being based on actual creativity of the sort prized by society. In summary of this brief on creativity studies, I would like to offer a few dogmatic-sounding observations. First, one need not be surprised at the fact that so-called creativity tests do not yield high correlations with IQ tests–but the correlations are generally far higher than those found in typical studies with range restrictions. I would anticipate that for normalized scores, the uncurtailed scatters for IQ versus creativity tests will be bivariate normal. Second, if we have honest to goodness criterion measures of literary or architectural or scientific creativity, the scatter diagram between IQ and such creativity (not normalized, since it makes sense to expect a skewed distribution for actual creativity) will be triangular in shape for unselected cases. That is, at the high IQ levels there will be a very wide range of creativity, whereas as we go down to average IQ, and on down to lower levels, the scatter for creativity will be less and less. Having a high IQ is not a guarantee of being creative; having a low IQ means creativity is impossible. Third, it remains to be seen whether or not the so-called creativity tests and/or factors derived therefrom have appreciable value as predictors of actual creative performance. Such tests may or may not yield better predictions than a test of general intelligence. Fourth, as far as I am concerned, to claim factorial validity for creativity tests, along with definitions of creativity in terms of tests, is an unwarranted avoidance of the fundamental problem of validity.

The recently renewed interest in 'gifted' children, along with the flurry of creativity studies, has led to a re-examination of methods for identifying the gifted. It has long been recognized that identification in terms of high

IQ is too narrow–those gifted in such areas as art and music would be overlooked. The argument against the IQ is now (Torrance, 1962) being reinforced by the claim that the selection of the top 20 per cent on IQ would mean the exclusion of 70 per cent of the top 20 per cent on tested creativity. This startling statistic, which implies a correlation of only 0.24 between IQ and creativity, is being used to advocate the use of creativity tests for identifying the gifted. Be it noted that these creativity tests will also miss those gifted in art and music.

We are being told that it is important 'to identify creative talent early in life', hence you need not be surprised that the search goes down to the kindergarten level, with claims of successful identification. The creativity tests are presumed to be better for this purpose than the IQ tests because of the failure of the IQ to be constant, an argument that completely overlooks the fact that the IQ does have some constancy whereas absolutely nothing is known about the stability of standings on creativity tests. The IQ tests, known to be imperfectly valid as predictors for outstanding achievement in life, are to be replaced by the creativity tests, known to be of unknown validity as predictors. Anyway, progress, defined as change, is in the offing.

The IQ is being linked with *learning* as an outmoded educational objective; the new objective involves an emphasis on *thinking*. Somehow or other, creativity, not general intelligence, is being associated with thinking. The horrible idea of underachievers and overachievers, in terms of expectancies based in the IQ, will be abolished. But no thought is given to the fact that the use of creativity tests will simply define a new crop of under- and overachievers.

In an apparent zeal to rid us of general intelligence, it is argued that measured creativity is significantly related to ordinary school achievement. Maybe so, but never, never does one find complete data reported as to the relative sizes of validity coefficients. And, as we have seen, the technique being used will show that equal coefficients are unequal. Why not the full facts, free of fantasy?

An additional difficulty is not being faced by those who would replace IQ tests by creativity tests, or creative-thinking tests. The factor analytic studies indicate either no, or a trivially small, general creativity factor in these tests, yet these self-characterized 'bold, adventurous' reformers (Torrance, 1963) do not hesitate to advocate a total score which is nearly devoid of meaning. Changing the curriculum to the teaching of creativity and creative thinking will not overcome this measurement difficulty. Again, I express the hope that the IQ is replaced by something better rather than by something worse.

There are other areas; such as reasoning, problem solving and concept formation, in which one might expect to find some consideration of intelligence as an aspect. One might also expect that investigators of thinking would have something to say about individual differences in thinking being dependent upon intelligence, but for some unintelligent reason these people seem never to mention intelligence. Surely, it cannot be inferred that thinking about thinking does not involve intelligence!

In conclusion

It has been the thesis of this paper that the concept of general intelligence, despite being maligned by a few, regarded as a second-order function by some, and discarded or ignored by others, still has a rightful place in the science of psychology and in the practical affairs of man. It has not been argued that the nature of general intelligence is well understood. Much, however, has been written about its nature. Over forty years ago, an editor secured and published the reasoned views of thirteen well-known test psychologists. Later, Spearman set forth his speculations about the nature of *g*. Prior to these, Binet had, of course, given much thought to the problem.

More recent discussions exist. Hebb (1949) has considered the problem from the viewpoint of neurology and brain functioning. Burt (1955), always a vociferous defender of the concept of general intelligence, has reviewed the evidence for a *g* and restated the idea, dreadful to some, that intelligence is innate. Perhaps it was inevitable that Cattell (1963), who has camped with the general intelligence contingent, should gaze into his crystal *n*-dimensional factor ball and find evidence for crystallized as opposed to fluid general intelligence. Hunt's (1961) book on *Intelligence and Experience* is in large part devoted to questions pertaining to the nature of intelligence.

By far the most provocative recent discussion that I have encountered is the closely reasoned 44-page paper by Hayes (1962). He puts forth a motivational-experiential theory of intelligence. In essence, he presumes that there are hereditary differences in motivation. 'Experience-producing drives' and environmental differences produce differences in experience, which in turn, by way of learning, lead to differences in ability. Therefore, differences called intellectual are nothing more than acquired abilities. I think that Hayes has ignored the possibility of individual differences in learning ability, but if such a formulation leads to experimental manipulation of variables, we may eventually make progress in an area that has too long been dominated by ever increasing fractionization by factor analysis, with little thought as to how the fractured parts get put together into a functioning whole.

Abilities, or capacities, or aptitudes, or intellectual skills, or whatever you choose to call them, are measured in terms of response products to standardized situations. The *stimulus* is presented to an *organism* which by some *process* comes up with a *response*; thus any attempt to theorize and/or study intellect in terms of a simple stimulus-response (S–R) paradigm seems doomed to failure unless drastically modified and complicated by the insertion of *O* for organism and *P* for process.

There have been thousands of researches on the multitudinous variations from organism to organism, and the results fill books on individual differences. These studies can be roughly classified into two types. First, those that ascertain the intercorrelations among scaled response products to various stimulus situations, known as tests, have to do with the structure of intellect; and whether the resulting factors are anything more than dimensions for describing individual differences need not concern us here. The second type of study seeks the non-test correlates of test performance, and whether or not any of the found correlates can be regarded as explaining individual differences is not of interest here. Both types of studies certainly force one to stress the overwhelming diversity exhibited among the organisms.

But these studies of individual differences never come to grips with the *process*, or operation, by which a given organism achieves an intellectual response. Indeed, it is difficult to see how the available individual difference data can be used even as a starting point for generating a theory as to the process nature of general intelligence or of any other specified ability.

As a basis for a little speculation, let us conceive of a highly hypothetical situation in which the two members of a pair of identical twins, with identical experiences, find themselves cast up on an uninhabited tropical island. Let us assume that they are at the super-genius level, far beyond that of your favourite man of genius. Let us also assume that, though highly educated in the sciences, they have been fortunate enough to have had zero exposure to psychology. In addition, we presume that, being highly involved and abstracted in the pursuit of science, they have never noticed what we call individual differences in abilities.

A quick exploration of the island assures them that food is plentiful, that shelter is available, and that clothing is not a necessity. To allay the boredom that they foresee as an eternity in this labourless heaven, they decide to spend their time in the further pursuit of science, but the lack of the wherewithal for constructing gadgets rules out any research in the physical sciences. Having had a college course in Bugs and Bites they proceed to study the life of the island's insects, then the habits of the birds, and the antics of a couple of monkeys. The manner in which the monkeys adjust to the environment leads them to set up some trial situations for more systematic observation. Needless to say, the monkeys show evidence of what we call learning and what we call problem solving.

Eventually they decide that attempting to outwit each other might be more fun than being outwitted by the monkeys, so they begin to cook up and use games and problems for this purpose. This activity leads each to speculate and introspect about how problems are invented and how solved. Then by cleverly designed experiments, preceded of course by theory, they set forth highly developed laws and principles about what we call reasoning and problem solving. Incidentally, they switch back and forth between roles of experimenter and subject, there being no college sophomores available.

They continue for years the study of their own mental operations, constantly on the alert for new phenomena to investigate.

And now with apologies to the ancient Greeks, who did have some ideas along these lines, we leave with you the 64-million drachma question: Will our two identical super-geniuses, being totally unaware of individual differences, ever hit upon and develop a concept of intelligence?

References

BARRON, F. (1963) *Creativity and Psychological Health.* New York: Van Nostrand.

BENNETT, G. K., SEASHORE, H. G. and WESMAN, A. G. (1952) *Differential Aptitude Tests, Manual* (2nd edn). New York: Psychological Corporation.

BURT, C. (1955) 'The evidence for the concept of intelligence', *Br. J. educ. Psychol., 25,* 158–77.

CATTELL, R. B. (1963) 'Theory of fluid and crystallized intelligence: a critical experiment', *J. educ. Psychol., 54,* 1–22.

CRONBACH, L. J. (1957) 'The two disciplines of scientific psychology', *Am. Psychol., 12,* 671–84.

GETZELS, J. W. and JACKSON, P. W. (1962) *Creativity and Intelligence.* New York: Wiley.

GHISELLI, E. E. (1963) 'Managerial talent', *Am. Psychol., 18,* 631–42.

GUILFORD, J. P. (1950) 'Creativity', *Am. Psychol., 5,* 444–54.

HARRELL, T. H. (1961) *Managers' Performance and Personality.* Cincinnatti: South Western.

HAYES, K. J. (1962) 'Genes, drives and intellect', *Psychol. Rep.,* 10, 299–342.

HEBB, D. O. (1949) *The Organization of Behaviour.* New York: Wiley.

HOLLAND, J. L. (1961) 'Creative and academic performance among talented adolescents, *J. educ. Psychol., 52,* 136–47.

HUNT, J. McV. (1961) *Intelligence and Experience.* New York: Ronald Press.

MacKINNON, D. W. (1962) 'The nature and nurture of creative talent', *Am. Psychol., 17,* 484–95.

MILES, T. R. (1957) 'On defining intelligence', *Br. J. educ. Psychol., 27,* 153–65.

SHAYCOFT, M. F., DAILEY, J. T., ORR, D. B., NEYMAN, C. A., Jnr. and SHERMAN, S. E. (1963) *Project Talent: Studies of a Complete Age-group–Age 15.* Pittsburgh: University of Pittsburgh.

SPEARMAN, C. (1904) '"General intelligence" objectively determined and measured', *Am. J. Psychol., 15,* 201–93.

TAYLOR, D. W. (1961) 'Variables related to creativity and productivity among men in two research laboratories'. In TAYLOR, C. W. and BARRON, F. (eds.) *Scientific Creativity: Its Recognition and Development.* New York: John Wiley.

THURSTONE, L. L. (1938) *Primary Mental Abilities.* Chicago: University of Chicago Press.

THURSTONE, L. L. and THURSTONE, T. G. (1941) *Factorial Studies of Intelligence.* Chicago: University of Chicago Press.

TORRANCE, E. P. (1962) *Guiding Creative Talent.* Englewood Cliffs, New Jersey: Prentice-Hall.

24 Race, intelligence and genetics: the differences are real

A.R. Jensen

In 1969, in the appropriately academic context of *The Harvard Educational Review* I questioned the then and still prevailing doctrine of racial genetic equality in intelligence. I proposed that the average difference in IQ scores between black and white people may be attributable as much to heredity as environment. Realizing that my views might be wrongly interpreted as conflicting with some of the most sacred beliefs of our democracy, I emphasized the important distinction between individual intelligence and the average intelligence of populations. Moreover, I presented my research in a careful and dispassionate manner, hoping that it would stimulate rational discussion of the issue as well as further research.

Much to my dismay, however, my article set off an emotional furor in the world of social science. Amplified by the popular press, the furor soon spread beyond the confines of academia. Almost overnight I became a *cause célèbre,* at least on college campuses. I had spoken what Joseph Alsop called 'the unspeakable'. To many Americans I had thought the unthinkable.

Science vs. the fear of racism

For the past three decades the scientific search for an explanation of the well-established black IQ deficit has been blocked largely, I feel, by fear and abhorrence of racism. In academic circles, doctrinaire theories of strictly environmental causation have predominated, with little or no attempt to test their validity rigorously. The environmentalists have refused to consider other possible causes, such as genetic factors. Research into possible genetic influence on intelligence has been academically and socially taboo. The orthodox environmental theories have been accepted not because they have stood up under proper scientific investigation, but because they harmonize so well with our democratic belief in human equality.

The civil-rights movement that gained momentum in the 1950s 'required' liberal academic adherence to the theory that the environment was responsible for any individual or racial behavioral differences, and the corollary belief in genetic equality in intelligence. Thus when I questioned such beliefs I, and my theories, quickly acquired the label 'racist.' I resent this label, and consider it unfair and inaccurate.

Source: *Psychology Today,* 1973, 7, 80-6.

The real meaning of racism

Since the horrors of Nazi Germany, and Hitler's persecution of the Jews in the name of his bizarre doctrine of Aryan supremacy, the well-deserved offensiveness of the term 'racism' has extended far beyond its legitimate meaning. To me, racism means discrimination among persons on the basis of their racial origins in granting or denying social, civil or political rights. Racism means the denial of equal opportunity in education or employment on the basis of color or national origin. Racism encourages the judging of persons not each according to his own qualities and abilities, but according to common stereotypes. This is the real meaning of racism. The scientific theory that there are genetically conditioned mental or behavioral differences between races cannot be called racist. It would be just as illogical to condemn the recognition of physical differences between races as racist.

When I published my article in 1969, many critics confused the purely empirical question of the genetic role in racial differences in mental abilities with the highly charged political-ideological issue of racism. Because of their confusion, they denounced my attempt to study the possible genetic causes of such differences. At the same time, the doctrinaire environmentalists, seeing their own position threatened by my inquiry, righteously and dogmatically scorned the genetic theory of intelligence.

Thankfully, the emotional furor that greeted my article has died down enough recently to permit sober and searching consideration of the true intent and substance of what I actually tried to say. Under fresh scrutiny stimulated by the controversy, many scientists have re-examined the environmentalist explanations of the black IQ deficit and found them to be inadequate. They simply do not fully account for the known facts, in the comprehensive and consistent manner we should expect of a scientific explanation.

The black IQ deficit

First of all, it is a known and uncontested fact that blacks in the United States score on average about one standard deviation below whites on most tests of intelligence. On the most commonly used IQ tests, this difference ranges from ten to twenty points, and averages about fifteen points. This means that only about 16 per cent of the black population exceeds the test performance of the average white on IQ tests. A similar difference of one standard deviation between blacks and whites holds true for eighty standardized mental tests on which published data exist.

A difference of one standard deviation can hardly be called inconsequential. Intelligence tests have more than proved themselves as valid predictors of scholastic performance and occupational attainment, and they predict equally well for blacks as for whites. Unpleasant as these

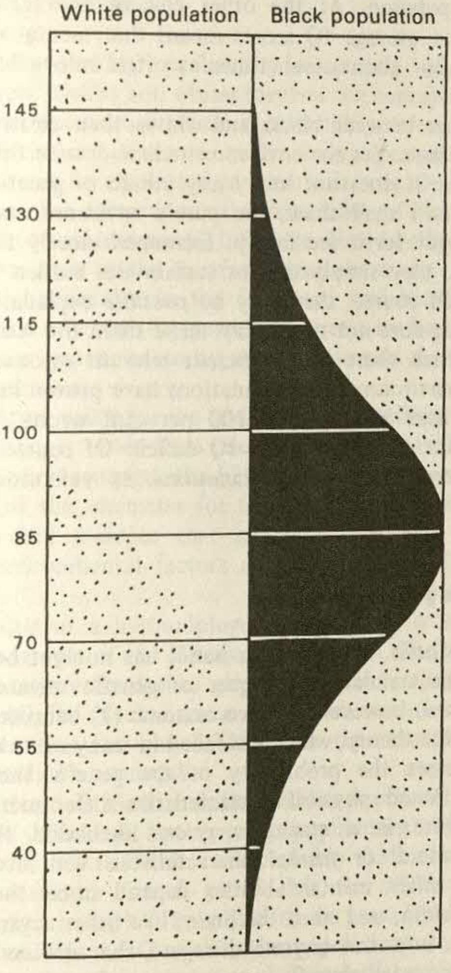


Fig. 1 Distribution of IQ scores. On most intelligence tests, black scores average about 15 points lower than do white scores

predictions may seem to some people, their significance cannot be wished away because of a belief in equality. Of course, an individual's success and self-fulfillment depends upon many characteristics *besides* intelligence, but IQ does represent an index, albeit an imperfect one, of the ability to compete in many walks of life. For example, many selective colleges require College Board test scores of 600 (equivalent to an IQ of 115) as a minimum for admission. An average IQ difference of one standard deviation between blacks and whites means that the white population will have about seven times the percentage of such potentially talented persons (i.e. IQs over 115) as the black population. At the other end of the scale, the fifteen-point difference in average IQ scores means that mental retardation (IQ below 70) will occur about seven times as often among blacks as among whites.

The IQ difference between blacks and whites, then, clearly has considerable social significance. Yet the environmentalists dismiss this difference as artificial, and claim it does not imply any innate or genetic difference in intelligence. But as I shall show, the purely environmental explanations most commonly put forth are faulty. Examined closely in terms of the available evidence, they simply do not sustain the burden of explanation that they claim. Of course, they may be *possible* explanations of the IQ difference, but that does not necessarily make them the *most probable*. In every case for which there was sufficient relevant evidence to put to a detailed test, the environmental explanations have proven inadequate. I am not saying they have been proven 100 per cent wrong, only that they do not account for *all* of the black IQ deficit. Of course, there may be other possible environmental explanations as yet unformulated and untested.

Arguments for the genetic hypothesis

The genetic hypothesis, on the other hand, has not yet been put to any direct tests by the standard techniques of genetic research. It must be seriously considered, however, for two reasons: (1) because the default of the environmentalist theory, which has failed in many of its most important predictions, increases the probability of the genetic theory; (2) since genetically conditioned physical characteristics differ markedly between racial groups, there is a strong *a priori* likelihood that genetically conditioned behavioral or mental characteristics will also differ. Since intelligence and other mental abilities depend upon the physiological structure of the brain, and since the brain, like other organs, is subject to genetic influence, how can anyone disregard the obvious probability of genetic influence on intelligence?

Let us consider some of the genetically conditioned characteristics that we already know to vary between major racial groups: body size and proportions, cranial size and shape; pigmentation of the hair, skin and eyes; hair form and distribution; number of vertebrae; fingerprints; bone density; basic-metabolic rate; sweating; consistency of ear wax; age of eruption of the permanent teeth; fissural patterns on the surfaces of the teeth; blood groups; chronic diseases; frequency of twinning; male-female birth ratio; visual and auditory acuity; colorblindness; taste; length of gestation period; physical maturity at birth. In view of so many genetically conditioned traits that do differ between races, wouldn't it be surprising if genetically conditioned mental traits were a major exception?

The heritability of intelligence

One argument for the high probability of genetic influence on the IQ difference between blacks and whites involves the concept of *heritability*. A technical term in quantitative genetics, heritability refers to the proportion of the total variation of some trait, among persons within a given population, that can be attributed to genetic factors. Once the heritability of that trait can be determined, the remainder of the variance can be attributed mainly to environmental influence. Now intelligence, as measured by standard tests such as the Stanford-Binet and many others, does show very substantial heritability in the European and North American Caucasian populations in which the necessary genetic studies have been done. I don't know of any geneticists today who have viewed the evidence and who dispute this conclusion.

No precise figure exists for the heritability of intelligence, since, like any population statistic, it varies from one study to another, depending on the particular population sampled, the IQ test used, and the method of genetic analysis. Most of the estimates for the heritability of intelligence in the populations studied indicate that genetic factors are about twice as important as environmental factors as a cause of IQ differences among individuals.

I do not know of a methodologically adequate determination of IQ heritability in a sample of the US black population. The few estimates that exist, though statistically weak, give little reason to suspect that the heritability of IQ for blacks, when adequately estimated, should differ appreciably from that for whites. Of course the absence of reliable data makes this a speculative assumption.

What implication does the heritability *within* a population have concerning the cause of the difference *between* two populations? The fact that IQ is highly heritable within the white and probably the black population does not by itself constitute formal proof that the difference between the populations is genetic, either in whole or in part. However, the fact of substantial heritability of IQ within the populations does increase the *a priori* probability that the population difference is partly attributable to genetic factors. Biologists generally agree that, almost without exception throughout nature, any genetically conditioned characteristic that varies among individuals within a subspecies (i.e. race) also varies genetically between different subspecies. Thus, the substantial heritability of IQ within the Caucasian and probably black populations makes it likely (but does not prove) that the black population's lower average IQ is caused at least in part by a genetic difference.

What about the purely cultural and environmental explanations of the IQ difference? The most common argument claims that IQ tests have a built-in cultural bias that discriminates against blacks and other poor minority groups. Those who hold this view criticize the tests as being based unfairly on the language, knowledge and cognitive skills of the white 'Anglo' middle class. They argue that blacks in the United States do not share in the same culture as whites, and therefore acquire different meanings to words, different knowledge, and a different set of intellectual skills.

Culture-fair vs. culture-biased

However commonly and fervently held, this claim that the black IQ deficit can be blamed on culture-biased or 'culture-loaded' tests does not stand up under rigorous study. First of all, the fact that a test is culture-*loaded* does not necessarily mean it is culture-*biased*. Of course, many tests do have questions of information, vocabulary and comprehension that clearly draw on experiences which could only be acquired by persons sharing a fairly common cultural background. Reputable tests, called 'culture-fair' tests, do exist, however. They use nonverbal, simple symbolic material common to a great many different cultures. Such tests measure the ability to generalize, to distinguish differences and similarities, to see relationships, and to solve problems. They test reasoning power rather than just specific bits of knowledge.

Surprisingly, blacks tend to perform relatively better on the more culture-loaded or verbal kinds of tests than on the culture-fair type. For example, on the widely used Wechsler Intelligence Scale, comprised of eleven different subtests, blacks do better on the culture-loaded subtests of vocabulary, general information, and verbal comprehension than on the nonverbal performance tests such as the block designs. Just the opposite is true for such minorities as Orientals, Mexican-Americans, Indians, and Puerto Ricans. It can hardly be claimed that culture-fair tests have a built-in bias in favor of white, Anglo, middle-class Americans when arctic Eskimos taking the same tests perform on a par with white, middle-class norms. My assistants and I have tested large numbers of Chinese children who score well above white norms on such tests, despite being recent immigrants from Hong Kong and Formosa, knowing little or no English, and having parents who hold low-level socioeconomic occupations. If the tests have a bias toward the white, Anglo, middle class, one might well wonder why Oriental children should outscore the white Anglos on whom the tests were originally standardized. Our tests of Mexican-Americans produced similar results. They do rather poorly on the culture-loaded types of tests based on verbal skills and knowledge, but they do better on the culture-fair tests. The same holds true for American Indians. All these minorities perform on the two types of tests much as one might expect from the culture-bias hypothesis. Only blacks, among the minorities we have tested, score in just the opposite manner.

Intelligence tests are colorblind

Those who talk of culture bias should also consider that all the standard mental tests I know of are colorblind, in that they show the same reliability and predictive validity for blacks and whites. In predicting scholastic achievement, for example, we have found that several different IQ tests predict equally well for blacks and whites. College-aptitude tests also predict grades equally well for blacks and whites. The same equality holds true for aptitude tests which predict job performance.

We have studied culture bias in some standard IQ tests by making internal analyses to see which kinds of test items produce greater differences in scores between blacks and whites. For example, we made such an item-by-item check of the highly culture-loaded Peabody Picture Vocabulary Test, on which blacks average some fifteen points lower than whites. The PPVT consists of 150 cards, each containing four pictures. The examiner names one of the pictures and the child points to the appropriate picture. The items follow the order of their difficulty, as measured by the percentage of the children in the normative sample who fail the item.

California vs. England; boys vs. girls

To illustrate the sensitivity of this test to cultural differences in word meanings, we compared the performance of white schoolchildren in England with children of the same age in California. Although the two groups obtained about the same total IQ score, the California group found some culture-loaded words such as 'bronco' and 'thermos' easy, while the London group found them difficult. The opposite occurred with words like 'pedestrian' or 'goblet'. Thus the difficulty of some items differed sharply depending on the child's cultural background. A similar 'cultural' bias shows up when comparing the performance of boys and girls, both black and white. Though boys and girls score about equally well over all, they show significant differences in the rank order of item difficulty; specific items, e.g. 'parachute' versus 'casserole', reflect different sexual biases in cultural knowledge.

Yet when we made exactly the same kind of comparison between blacks and whites in the same city in California and even in the same schools, we found virtually no difference between the two groups in the order of items when ranked for difficulty, as indexed by the per cent failing each item. Both groups show the same rank order of difficulty, although on each item a smaller percentage of blacks give the correct answer. In fact, even the differences between adjacent test items, in terms of percent answering correctly, show great similarity in both the black and white groups.

If this kind of internal analysis reflects cultural bias between different national groups, and sexual bias *within* the same racial group, why does it not reflect the supposed bias *between* the two racial groups? If the tests discriminate against blacks, why do blacks and whites make errors on the same items? Why should the most and least common errors in one group be the same as in the other?

Another way internal analysis can be used to check for bias involves looking for different patterns of item intercorrelations. For example, if a person gets item number 20 right, he may be more likely to get, say, item 30 right than if he had missed item 20. This follows because the test items correlate with one another to varying degrees, and the amount of correlation and the pattern of intercorrelations should be sensitive to group differences in cultural background. Yet we have found no significant or appreciable differences between item intercorrelations for blacks and whites.

In summary, we have found no discriminant features of test items that can statistically separate the test records of blacks and whites any better than chance, when the records are equated for total number correct. We could do so with the London versus California groups, or for sex differences within the same racial group. Thus, even when using the PPVT, one of the most culture-loaded tests, black and white performances did not differ as one should expect if we accept the culture-bias explanation for the black IQ deficit. I consider this strong evidence against the validity of that explanation.

The effect of the tester

What about subtle influences in the test situation itself which could have a depressing effect on black performance? It has been suggested, for example, that a white examiner might emotionally inhibit the performance of black children in a test situation. Most of the studies that have attempted to test this hypothesis have produced no substantiation of it. In my own study in which 9,000 black and white children took a number of standard mental and scholastic tests given by black and white examiners, there were no systematic differences in scores according to the race of the examiners. What about the examiner's language, dialect, or accent? In one study, the Stanford-Binet test, a highly verbal and individually administered exam, was translated into black ghetto dialect, and administered by a black examiner fluent in that dialect. A group of black children who took the test under these conditions obtained an average IQ score less than one point higher than the average IQ score of a control group given the test in standard English.

The question of 'verbal deprivation'

To test the popular notion that blacks do poorly on IQ tests because they are 'verbally deprived,' we have looked at studies of the test performances of the most verbally deprived individuals we know of: children born totally deaf. These children do score considerably below average on verbal tests, as expected. But they perform completely up to par on the nonverbal culture-fair type of tests. Their performance, then, turns out to be just the opposite of the supposedly verbally deprived blacks, who score higher on the verbal than on the nonverbal tests.

If one hypothesizes that the black IQ deficit may be due to poor motivation or uncooperative attitudes of blacks in the test situation, then one must explain why little or no difference in scores occurs between blacks and whites on tests involving rote learning and memory. Such tests are just as demanding in terms of attention, effort and persistence, but do not call upon the kinds of abstract reasoning abilities that characterize the culture-fair intelligence tests. We have devised experimental tests, which look to pupils like any other tests, that minimize the need for reasoning and abstract ability, and maximize the role of nonconceptual learning and memory. On these tests black and white children average about the same scores. Therefore, the racial difference clearly does not involve all mental abilities equally. It involves mainly conceptual and abstract reasoning, and not learning and memory.

Another factor often cited as a possible explanation for the black IQ deficit is teacher expectancy—the notion that a child's test score tends to reflect the level of performance expected by his or her teacher, with the teacher's expectation often based on prejudice or stereotypes. Yet numerous studies of teacher expectation have failed to establish this phenomenon as a contributing factor to the lower IQ scores of blacks.

Testing the environmental hypothesis

To test the environmentalist hypothesis, we have examined the results of those tests that most strongly reflect environmental sources of variance, and they turn out to be the very tests that show the least difference between blacks and whites in average scores. The greatest difference in scores between the two racial groups occurs on the tests we infer to be more strongly reflective of genetic variance. If the cultural-environmental hypothesis were correct, just the opposite would be true.

The 'sociologist's fallacy'

In an attempt to disprove the genetic hypothesis for the black IQ deficit, environmentalists frequently cite studies that compare IQs of matched racial groups, and find considerably less difference in test scores than the usual fifteen-point difference between races. Here we have a good example of the 'sociologist's fallacy'. Since whites and blacks differ in average socioeconomic status (SES), the matching of racial groups on SES variables such as education, occupation, and social class necessarily means that the black group is more highly selected in terms of whatever other traits and abilities correlate with SES including intelligence. Therefore the two groups have been unfairly matched in terms of IQ.

Those who cite the socioeconomic matching studies also fail to take account of the well-established genetic difference between social classes, which invalidates their comparison. For example, when the two races are matched for social background, the average skin color of the black group runs lighter in the higher SES groups. This difference indicates that genetic characteristics do vary with SES. Thus, SES matching of blacks and whites reduces the IQ difference not only because it controls for environmental differences, but because it tends to equalize genetic factors as well.

Variables that don't behave

A host of other environmental variables don't behave as they ought to according to a strictly environmentalist theory of the black IQ deficit. For example, on practically all the socioeconomic, educational, nutritional and other health factors that sociologists point to as causes of the black-white differences in IQ and scholastic achievement, the American Indian population ranks about as far below black standards as blacks do below those of whites. The relevance of these environmental indices can be shown by the fact that within each ethnic group they correlate to some extent in the expected direction with tests of intelligence and scholastic achievement. Since health, parental education, employment, family income, and a number of more subtle environmental factors that have been studied are all deemed important for children's scholastic success, the stark deprivation of the Indian minority, even by black standards, ought to be reflected in a comparison of the intelligence and achievement-test performance of Indians and blacks. But in a nationwide survey reported in the Coleman Report, in 1966, Indians scored *higher* than blacks on all such tests, from the first to the twelfth grade. On a nonverbal test given in the first grade, for example, before schooling could have had much impact, Indian children exceeded the mean score of blacks by the equivalent of fourteen IQ points. Similar findings occur with Mexican-Americans, who rate below blacks on socioeconomic and other environmental indices, but score considerably higher on IQ tests, especially on the nonverbal type. Thus the IQ difference between Indians and blacks, and between Mexican-Americans and blacks, turns out opposite to what one would predict from purely environmental theory, which of course, assumes complete genetic equality for intelligence. No testable environmental hypothesis has as yet been offered to account for these findings.

Does malnutrition affect intelligence?

What about malnutrition, another factor frequently cited by the environmentalists to disprove the genetic hypothesis? Malnutrition has indeed been found to affect both physical and mental development in a small percentage of children in those areas of the world that suffer severe protein deficiencies: India, South America, South Africa and Mexico. But few blacks in the US show any history or signs of severe malnutrition, and I have found no evidence that the degree of malnutrition associated with retarded mental development afflicts any major segment of the US population.

Nor do I know of any evidence among humans that maternal malnutrition, by itself, can have pre- or postnatal effects on a child's mental development. The severe famine in the Netherlands during the last years of World War II provided an excellent case study of such a possibility. Thousands of men conceived, gestated, and born during the period of most severe famine, were later tested, as young adults, on Raven's Standard Progressive Matrices, a nonverbal reasoning test. Their scores did not differ significantly from the scores of other Dutch youths of the same age who had not been exposed to such maternal nutritional deprivation.

If further research should definitely establish the existence of genetically conditioned differences in intelligence between certain races, what would be the practical implications? It would take several articles to consider the question adequately, but the only morally tenable position in human relations would remain unchanged: that all persons should be treated according to their own individual characteristics, and not in terms of their group identity. Let me stress that none of the research I have discussed here allows one to conclude anything about the intelligence of any individual black or white person.

Equality of rights and opportunities is clearly the most beneficial condition for any society. Acceptance of the reality of human differences in mental abilities would simply underline the need for equality of opportunity in order to allow everyone to achieve his or her own self-fulfillment. In order to take account and advantage of the diversity of abilities in the population, and truly to serve all citizens equally, the public schools should move beyond narrow conceptions of scholastic achievement. They should offer a much greater diversity of ways for children of whatever aptitude to benefit from their education.

Environment vs. genetics: still an open question

I have tried to emphasize the uncertainty of our knowledge of the causes of race differences in mental abilities. I do not claim any direct or definite evidence, in terms of genetic research, for the existence of genotypic intelligence differences between races or other human population groups. I have not urged acceptance of a hypothesis on the basis of insufficient evidence. I have tried to show that the evidence we now have does not support the environmentalist theory, which, until quite recently, has been accepted as scientifically established. Social scientists have generally accepted it without question, and most scientists in other fields have given silent assent. I have assembled evidence which, I believe, makes such complacent assent no longer possible, and reveals the issue as an open question, calling for much further scientific study.

Politicizing a scientific issue

Most of the scientists and intellectuals with whom I have discussed these matters in the past few years see no danger in furthering our knowledge of the genetic basis of racial differences in mental or behaviorial traits. Nor do they fear general recognition of genetic differences in such traits by the scientific world, if that should be the eventual outcome of further research. They do see a danger in politicizing a basically scientific question, one that should be settled strictly on the basis of evidence.

Most of the attempts to politicize the issue, I have found, come from the radical left. True liberals and humanists, on the other hand, want to learn the facts. They do not wish to expend their energies sustaining myths and illusions. They wish to face reality, whatever it may be, because only on the level of reality can real problems be effectively confronted. This means asking hard questions, and seeking the answers with as much scientific ingenuity and integrity as we can muster. It means examining all reasonable hypotheses, including unpopular ones. It means maintaining the capacity to doubt what we might most want to believe, acknowledging the uncertainties at the edge of knowledge, and viewing new findings in terms of shifting probabilities rather than as absolute conclusions.

25 Environment: the cumulation of effects is yet to be understood

A.L. Stinchcombe

This essay is divided into two broad parts. First I want to make several comments about what Jensen's (1969) evidence shows, taking the evidence at face value. In general, Jensen conceives the evidence he presents to have a single obvious interpretation. I think the singularity of interpretation is often the result of a kind of systematic naivete about how the environment works.

Second, I want to develop the outlines of a theory about how the environment determines the abstract cognitive structures people have to deal with the environment–especially the symbolic environment–which are measured by IQ tests. The basic argument will be that many of the findings in Jensen's article fit into a pattern that would be produced if rural and oppressed cultures were inefficient in teaching advanced cognitive structures at later ages. Jensen assumes that if environment works harm, it will do so in the learning of simple cognitive skills at earlier ages. This might be called the 'Headstart' environmental theory. When he gives arguments against this theory, he thinks he is arguing for a genetic theory. But he is also implicitly arguing for a 'civilization' theory of environmental determination of thought, which says that cultures differ most at the most sophisticated levels. My purpose will be to use his evidence in favor of a different theory of environmental impact on IQ.

What do the kinship data show ?

The most impressive argument for the heritability of intelligence is clearly the data on correlations among kinsmen in IQ. The basic finding is that you take pairs of kinsmen that are genetically more closely related (e.g. identical twins) and pairs of kinsmen that are farther apart (e.g. fraternal twins), then the correlation between the IQs of the closer pairs will be greater than the correlation between the IQs of the more distant pairs. That is, if you know the IQ of one identical twin, you can predict the IQ of the other very well. If you know the IQ of a fraternal twin, you cannot predict the IQ of the other nearly as well. By combining genetic theory about *differences in genetic closeness* of different types of relatives with knowledge of the *differences in correlations* between pairs of different types of relatives, one can estimate how much effect genetic closeness has. From this estimate of

Source: *Harvard Educational Review*, 1969, *39*, 511-22.

the effect of genetic closeness, one can estimate the *size* of the effect of genetic constitution, even though one cannot measure genetic constitution itself. This seems to me a valid procedure for what it is designed for, though subject to the risks of any attempt to get at unobserved variables from the pattern of manifest variables.[1]

But if we look at Jensen's estimates of environmental effects for IQ, we find that the difference *in environment* between one twin and another is the same size as the difference *in environment* between one family and another. That is, if you pick out two brothers at random, the chance that one of them will be in a 'deprived environment' while the other is in an 'advantaged environment' is the same as the chance that one out of a pair of randomly picked families is a 'deprived environment' while the other is advantaged. This clearly is not the notion most people have of variations in environment.

This peculiar fact of as much environmental variation within families as between families is illuminated if we look at the samples. The one described for London is more or less typical. To summarize briefly what is going on, if you and your cousin both live in London and both are still in school, then the difference between your father's family and your uncle's family is about of the same order of magnitude as the difference between your family as you grew up in it and your family as your brother grew up in it. That is, the environmental variations we are talking about are variations *within* an endogamous group, a group that intermarry and exchange children by adoption. If we want to generalize across endogamous groups (groups that intermarry among themselves), for instance across races, we have to ask whether there is more variation in environment between a white and Negro family than there is among kinfolk living in London. If there is, then the between-families variance in environmental factors would increase, and the relative role of genetic factors would decrease. Also, the relative role of within-family variation in environment would decrease relative to between-family variation.

What is really needed to apply such kinship research to the problem of racial differences are pairs of identical twins, one of whom is a Negro, and pairs of fraternal twins, one of whom is a Negro.

Regression effects

Another apparently impressive piece of evidence is the larger regression effect for Negroes rather than whites. Briefly the data are mean IQs of children, subtracted from the mean IQs of parents in each of three, four, or five social classes. (Roughly speaking this is what is going on—the actual calculations do not seem to be as adequate to the purpose as I described.) The differences are larger for Negro middle-class children than for white middle-class kids. The argument here goes as follows. The difference between childrens' and parents' IQ measures the difference between the particular set of genes that are manifested in the parent and the average of the total set of genes which the parent carries. That is, out of the total pool of genes that the parent carries only one subset are the ones that he gives his children. If the manifested subset is very different from the pool, then the children will be very different from their parents. So the size of the difference between parents and children is supposed to measure the distance *for the parents* between the genes they manifest and the genes they carry without manifesting. If Negroes of high intelligence show more regression effect than whites of the same intelligence and environment, that is supposed to show that the pool of genes from which exceptionally talented Negroes come is farther from them, on the average, than the pool from which exceptionally tallented whites come.

---

[1] One assumption it does make is that if two people are born in the same family they have the same father. This is, of course, a theory of the frequency of extramarital affairs or of extramarital contraception. It has been estimated that a fourth of all married women in the US have extra-marital affairs, but I know of no studies of their contraceptive practice in such affairs. One could estimate the frequency of such fruitful affairs by studying some clearly inherited characteristics, such as blood types.

Now the key to this inference is actually equating the environments of the children. For if the children of talented Negroes are actually exposed to a less cognitively rich environment than are whites, even though they are in the 'same' social class as measured, then we would expect a larger regression effect even if the pool of genes were the same distance from the parents. What we have to do then is to examine the equation of parents' status first, and then whether the equation of status equates environments. Briefly the argument will be that Jensen has not in fact equated statuses, and that even if he had he would not have equated environments.

First, let us assume that the indicator of status in the studies that Jensen quotes were a perfect measure of status. Then what he has done is to pick out three (or four, or five) cutting points on a continuous variable, and equated all people above the top cutting point as middle class, and so forth. Now as Jensen so carefully shows for IQ, if one distribution has a lower mean than the other and the distributions have the same shape, the more extreme values are more under represented on the lower distribution than the more moderate values. The same is true of status. Rockefellers, Mellons, and du Ponts are more underrepresented among Negroes than are $20,000 a year men; $20,000 a year men are more underrepresented than $10,000 a year men, and so forth. So if you take a given cutoff point for middle class, those few Negroes that are middle class will be mostly very near the cutoff. Many of the whites will be quite far from the cutoff. Thus actually the mean class position for middle-class Negroes is likely to be much lower than the mean class position of middle-class whites. Thus even if the class measures were perfect, the groups would not be equated.

But of course the actual situation is that the measurement of social status is much more unreliable than the measurement of IQ, and Jensen is always careful to take account of measurement error for IQ. Merely to take occupation as an example, at a given point in time when the census reinterviews people about their occupations, they get 17 to 22 per cent giving different occupation (Blau and Duncan, 1967). Of course the census uses trained interviewers who know what is required to classify an occupation. Most school studies ask children to describe their parents' occupation. Census studies of the reliability of occupational scales recently reported give figures in eighties, somewhat lower for Negroes than for whites. Questionnaires, especially from children, must have a great deal more measurement error than census interviewers.

Furthermore, the environment of a child is a cumulative matter, consisting of so and so many months in a middle-class family when his father had a good job, so and so many in a welfare family when he was unemployed, and so forth. Thus even if we had a perfect measure of the status of a family at a given time, it would be a poor measure of the cumulative environment of the child if peoples' status changes much over their lives.

Perhaps this will be more understandable with an example. Suppose one truly has 10 per cent of middle-class Negroes and 90 per cent working class, while whites are 50 per cent middle class and 50 per cent working class. And, also, assume that the likelihood of misclassification is the same for both races and both social classes. Now apply a measurement that misclassifies 10 per cent of all people. Applying this to Negroes, we get:

		Classified as Middle Class
90% of the 10% Truly Middle Class	=	9%
10% of the 90% Truly Working Class	=	9%
Total classified as Middle Class	=	18%

Percentage of Negroes Measured Middle Class Who are Truly Workers = 50%

Applying the same measurements to whites, we would get:

90% of the 50% Truly Middle Class	=	45%
10% of the 50% Truly Working Class	=	5%
Total classified as Middle Class	=	50%

Percentage of Whites Measured Middle Class who are Truly Workers = 10%

Thus we would expect that many more of the Negroes Jensen calls middle class are truly working class (either at present, or in the cumulative environmental sense) than are the whites. Hence because of measurement error he has not equated the environments. Actually in most of the studies he reports on, the most important environmental variable, the IQ of the parent, has not been equated at all.

Finally, of course, children associate with other children as well as with their parents. With a highly segregated society, this means that Negro children associate with workers' children, while white children associate with middle-class children. Consequently even were the status of families equated, the status of interpersonal environments of children would not be equated.

On all these grounds, then, the different regression effects cannot be taken as evidence that talented Negroes are farther genetically from their gene pools than are talented whites.

How big is a big environmental effect?

The difficulty Jensen has in thinking about environment as cumulative shows up again in his evaluation of the size of environmental effects. Jensen would like to have environments change once and for all from one environment to another, as genes do, and then stay that way. We might call this the spurt conception of the environment. Environmental change in this conception comes in a spurt, and any time after that we can measurer the effects of the spurt. Genetic variation does come in spurts; environmental variation does not.

The difference between Negro and white measured IQ scores in the US is about one standard deviation, or fifteen IQ points. Jensen observes that by decreasing the fear people have in the testing situation, so that they feel they control the situation and can keep from getting hurt, one can often increase poor children's IQs by five to eight points, or about one-third to one-half of the distance between whites and Negroes. Now this seems to him to be evidence of a small environmental effect, since most *other* changes in environments only produce about the same effect. But this is only because he regards taking fear out of social situations as eliminating measurement error. If he regarded it as changing the environment, he would conclude that taking fear out of the relations of Negroes to their environment might, by itself, decrease the difference between Negro and white performance.

Then Jensen goes through a set of studies which show that by changing the environment of children for six, or nine, or three months, one can change their IQs by five to ten points. That is, a short period of environmental enrichment can apparently wipe out a third of the difference between whites and Negroes. Furthermore, this is change in only 25 or 30 hours a week of their environment.

Finally, we quite often find that naturally occurring variations in environments cause variations of this order of magnitude in IQ or mental ability scores. Jensen himself quotes a study which estimates that a standard deviation of environmental variation within a white group can cause two-fifths of a standard deviation in IQ. To explain the variation between Negro and white IQ on this environmental basis, we need to find determinants of IQ on which there is a difference of two and one-half standard deviations between Negroes and whites, assuming a linear relationship between the variables. This is a difference such that, for a normally distributed variable, about one per cent of Negroes would be above average of the whites. It does not seem to me outrageous that a well-measured variable of oppressiveness of conditions of life and cultural deprivation might show such a difference.

Besides this estimate of Jensen's, we can get a notion of the size of the environmental variation by observing variation between Negroes whose environment approximates that of whites from those whose environment approximates that of Negroes. Clearly if Negroes go to nearly all-white schools, they are exposed for thirty hours a week to the same environment most whites are. If they go to all Negro schools, the thirty hours a week are like those to which Negroes are exposed. Coleman's (1966) estimates of the correlation between proportion white and Negro mental abilities indicates that by changing this part of a child's environment one makes about a half a standard deviation difference. This is of course also a non-cumulated score. A Negro child in an all-white school is much more likely than a white child in the same school to have spent part of his life in an all-Negro school.

The estimate of the effect of variations among white families in the kinship studies comes to approximately the same conclusion. If the correlation between adopted children and their parents in IQ is taken as a measure of the effect of environment, environmental variation between families who adopt each others' children explains about 24 per cent of the variance. That means that a standard deviation of environmental variation in families that adopt each others' children would cause about a half standard deviation variation of IQ.

What we have then is a large number of environmental variables, many of them occupying a small portion of children's lives, that explain between one-third and two-thirds of the difference between races. Equalize the amount of fear between races, put Negroes in all-white schools, or schools with the same conditions, improve the standards of Negro families by a standard deviation, then (if the effects are additive) the IQs would be equalized. Jensen thinks that each such variation is small because it only explains half or two-thirds of the racial difference. But environments cumulate.

Because Jensen thinks of environment in spurts rather than in hours per day, he is 'surprised' when an equalizing spurt does not equalize IQ. He translates this surprise into an argument that the spurt that created the genetic differences is more important. Suppose that we have a difference in environment which gives one group a rate of improvement in cognitive skills of 5 per cent a year, while another group advances at 3 per cent a year. The ratio of advance would be 1.05/1.03, or about 1.02 per year. If the two groups start equal, the more advantaged group will be about two per cent ahead after a year. If this environmental difference is maintained for, say, the first twenty years of life, the cumulated effect is about $(1.02)^{20}$, or about 1.5. The effect of a given year's enriched environment, which equalizes the increase for one year, is about a fifth of the disadvantage at five years of age. But it is about one-twenty-fifth of the disadvantage at twenty years of age. Jensen seems surprised that equalizing a single year results in an effect which disappears in a few more years. He would not be so surprised if he had thought a bit about a compound interest table.

The cognitive environment of the individual

So far I have let Jensen get away with his conception of intelligence, and of the environment, because most of his arguments can be undermined even if we allow him his conception of his own business, and require only elementary elaboration of his conception and measurement of environment. It seems to me that the rest of his argument stands up at its own level, and requires some more fundamental thought about cognition and the sociology of cognition. What I will try to do now is to develop a theory of what intelligence, as measured by the usual IQ tests, is, and then on this basis develop a notion of what kind of environmental variables might determine it. Then I will try to derive the remainder of Jensen's results from this alternative theory. The combination of his genetic spurt conception of intelligence and his Headstart theory of environment makes it hard for him to think about cognitive functioning appropriately.

First, to intelligence. The central fact about intelligence among children is that older children have to be able to do *more* than younger children in order to stay at the same level of intelligence. Jensen notes that Negro children's capacities go down from a slight superiority to whites at early ages to inferiority at later ages. That does not mean, of course, that Negro children forget by five years old what they knew how to do at three. It means instead that the time between mastery of one task or test and mastery of a more complex task or test is longer for Negro children than for white children.

That is, from the point of view of an individual child's biography, his intelligence is a measure of the time between being able to manage one type of task and being able to manage the next. Thus tasks are arranged more or less in layers with one group of tasks or intelligence test items being a layer which is normally learned after the 'simpler' ones and before the more 'complex' ones.

What seems to be behind each layer of tasks is one or more *cognitive structures,* or styles of abstraction and reasoning. Once a child learns to manage a given general cognitive structure or style of abstraction and reasoning, he can usually solve or answer most of the tasks or test items using that cognitive structure. The cognitive structure may involve perception, operations on perceptions, combinations of perceptions with previous knowledge, and so forth. There is a fair amount of evidence that such structures are well ordered, so that a five-year-old average IQ score is rarely made up of a hash of questions at a four-year-old level and a fifteen-year old level. Instead the normal five-year-old answers four-year-old and five-year-old questions. The six-year-old answers four, five, and six-year-old, etc. The classic work here was done by Jean Piaget. The funny scatter diagram in Jensen seems to be evidence in Piaget's favor. (The subtitle of these diagrams is mixed up with something else.) If we think of intelligence as such a layer-cake type of phenomenon in the individual, rather than as a manifestation at the present time of his constant IQ 'dispositional property', then we can ask what sorts of environmental features would be most likely to affect it. Probably the most important characteristic of the environment is the frequency with which the *next developmental type* of cognitive structure is used in a child's environment. Most of us at breakfast use grunts and groans and expressions of displeasure whose cognitive content is easily mastered, if not enjoyed, by a three-year-old. During the day we vary in intellectual level, returning just before sleep to the same level of grunts and groans, if we are lucky. The same variation is true of a child's playmates–they mostly function below their capacities, and only occasionally reach their limits. Although Jensen sometimes seems to subscribe implicitly to the Education School Fallacy (that anyone is smart enough to teach kids something),[2] the same variation in intellectual level is characteristic of classrooms.

What we would like, then, as an environmental measure of development pressure, is a measure of the density of interaction of the child with someone who, in the interaction, leads him up one cognitive structural level from where he is, solves his problem, then restates it at the level where the child is. (Compare the account of language learning in Brown (1965) *Social Psychology*.) Since most people most of the time function with structures much less complex than they manage in their finest hour, a person's cognitive behavior is probably distributed with different intensities at different levels below his highest capacities. These intensities or densities are probably determined by various influences, such as the intelligence of people around him, the complexity of the problems in his work, his amount of education, and so forth, as well as being limited by his upper limit of cognitive structures (his IQ). It is well known, for example, that professional men's IQs do not decline with age as rapidly as do the IQs of manual workers. Presumably this is because they keep the highest levels of mental functioning in better shape by more practice.

---

[2] Jensen, for instance, wonders why experimental programs, when made mass programs, very often have less effect. Perhaps the mass of teachers are dumber than those involved in experimental programs.

Thus there should be two main causal forces which would determine the mean level and variations of the cognitive functioning of a man–his *capacity to abstract* (his IQ) and his socially and psychologically induced *inclination to abstract* (determined by his education, his occupation, the intelligence of his wife and friends, and their educations and occupations, etc.). Thus we would expect to find correlations between people's IQ and the rate of development of children with whom they come in contact. For instance, there should be correlations between the IQ of foster children and the IQ of their foster parents (reported by Jensen, 1969), between the IQ of teachers and the achievement of children in their schools (reported by Coleman *et al.*, (1966)–the effect is stronger at higher grades), between the IQ of peers and the rate of advance of the IQ of peers (difficult to isolate, but fairly clear in some Coleman data).

Given the proportion of a person's cognitive functioning which just strains the child, the next question is the *hours of attention* he gives to the child, and the *attentiveness of the child* during those hours. We have very few estimates of these things, but we suppose that mothers spend more time with children than fathers (hence there should be higher correlations with mothers' IQ), that parents spend more time with only children than with many children (hence only children should be smarter generally and have less regression from parents' IQ), and so forth. The hours of attention received by and paid by children at different levels of cognitive tasks are ordinarily not measured in educational experiments. If the change from small to large programs changes the level of attentiveness of teachers, or students, or both, then we would expect less change in cognitive functioning in the expanded programs.

Civilization and intelligence

We could now define the level of civilization of a group by the average level of cognitive function used to solve its problems. That is, an 'advanced' civilization would be one whose average man at the average moment was using, say, a third grade level of abstraction. There would be a wide distribution around that level, of course. A 'backward' people would be one which did not ordinarily use the cognitive structures (e.g. mathematical reasoning, experiments, etc.) of which it is genetically capable. In general, more complex cognitive processes require reading and writing. Few people can solve mathematical problems in their heads that they solve easily on paper. Probably a good surrogate indicator of the level of cognitive functioning of a group is hours per day spent in reading and writing.

Ever since a famous sociologist of cognitive affairs in 1848 spoke of 'the idiocy of rural life', evidence has been accumulating that farm life has not provided the practice in symbolic analysis of the environment that a city did. Civilizations grow in cities. In some rich agricultural areas (e.g. Iowa, New

Zealand) civilization has more or less completely spread to the countryside.

If we suppose that the definition given above of civilization is something like the same thing that we ordinarily call civilization, then it will come as no surprize to find that rural people when tested have regularly turned up less intelligent on IQ tests. Furthermore groups imported for plantation labor and not allowed to develop their own civilized institutions (schools, churches, publishing houses, newspapers, political parties, trade unions, welfare groups, community governments, etc.), show the most IQ disadvantage compared to urban people. They also show the most gains in IQ when moved to the more civilized environment of an urban slum. They show still further gains as the hours per day of reading and writing increase in the group.

What this suggests then is that one of the legacies of slavery in the Negro community is a dearth of institutions that function routinely at advanced cognitive levels. There are too few Negro colleges, too few professionals, newspapers and book publishers, and so forth, to fill environment with a rich variety of cognitive levels. But it might be that as Negro children grow older, after starting to learn with the same capacity as white children, their environment is progressively less rich in the frequency of use of the next developmental cognitive structure.

If this were the case, we would expect that Negro IQ would appear to decline, as it took them longer to learn each successive cognitive structure. This is what Jensen reports. We might also expect there to be stability of IQ disadvantage, even though this looks as if it might be evidence for genetic determination of IQ.[3]

The purpose of these last two sections has been to develop a more serious environmental theory of intelligence than the Headstart theory, to suggest how research might be done on such a theory, and to explain why we might expect on the basis of such a theory that an oppressed rural group from a backward section of the country might show a progressive deficit. Instead of defining environment in a way borrowed from other fields, an environmental theory of determinants of cognitive capacity needs to define the environment in intellectual terms.

In particular, we need to know the proportion of time a child focusses attention on cognitive structures at the next level of development above his own. Various results that we already have can be interpreted in the light of such a conception of the environment. For instance, the correlation of children's achievement with the mean vocabulary scores of their teachers makes sense if we think that smarter teachers will strain their students' capacities a little more.

---

[3] An IQ deficit of 15 points, or one standard deviation, is approximately equivalent to the amount of development that an average child has gone through in the last sixth of his life. That is, for a child of six, an IQ of 85 means he can do approximately the things a normal 5-year-old can. For a child of 15, it means he can do approximately the things a normal 12-year-old can. Thus a constant deficit of one standard deviation means functioning at progressively more years behind others.

The idiocy of rural life, the cognitive consequences of growing up in urban slums, the disparities of achievement among ethnic, religious, and social-class groups begin to make sense if we define civilization in terms of the densities or frequencies of use of intellectual structures at different levels.

References

BLAU, P.M. and DUNCAN, O.D. (1967) *The American Occupational Structure.* New York: John Wiley and Sons.

BROWN, R. (1965) *Social Psychology.* New York: The Free Press.

COLEMAN, J.S., *et al.* (1966) *Equality of Educational Opportunity.* Washington: US Government Printing Office.

JENSEN, A. R. (1969) 'How much can we boost IQ and scholastic achievement?', *Harvard Educational Review, 39,* 1-123.

26 The person, the product and the response: conceptual problems in the assessment of creativity

P.W. Jackson and S. Messick

The ultimate concern of the psychologist is the human mind and its inventions. Although there are many ways to describe man's mental complexity–and particularly to depict his cognitive strengths–the two terms 'intelligence' and 'creativity' seem to have the greatest summary power. It is this concentration of meaning that explains the endurance of these two words in the layman's language and their continued use in professional discussions.

Typically, efforts to distinguish empirically between the concepts of creativity and intelligence have concentrated on attempts to demonstrate that performance on tests requiring the production of unusual and remote responses is independent (to some extent) of performance on conventional tests of intelligence. But as Golann quite correctly points out in his recent review of studies of creativity,

> . . . these data are in a sense arbitrary: intelligence is not performance on a test; creativity is more than test performance or being judged as creative. What is needed for the understanding of the relationship between creativity and intelligence is not only data at the correlational level, but conceptual reorganisation as well (Golann, 1963).

The controversy created by some of the recent research studies of creativity seems to have focused attention not on unresolved conceptual problems but on the empirical specification of particular, and sometimes arbitrary, aspects of the creativity-intelligence distinction. Moreover, the interesting results obtained with tests of fluency and remote associations have diverted our attention temporarily from the important task of considering other criteria, in addition to the production of unusual responses, essential to the total assessment of creativity.

The evaluation of intellectual performance

The exercise of a person's cognitive powers is often the occasion for evaluation. People are continually being informed that they have done well or poorly on intellectual tasks, and the variety of forms these judgments

---

Source: *Journal of Personality*, 1965, *33*, 1-19.

may take is almost as great as the variety of behaviours being judged. The wide assortment of evaluative comments can be crudely classified, however, into two overlapping categories. On the one hand, there are those judgments having to do with the 'correctness' or 'rightness' of a person's response. These evaluations deal with the degree to which certain objective and logical criteria have been satisfied. On the other hand, there are judgments having to do with the worth or 'goodness' of a person's response. These evaluations deal with the degree to which certain subjective and psychological criteria have been satisfied. Although, as has been suggested, these two classes of evaluations overlap (and some kinds of behaviour satisfy both sets of criteria), the two are nonetheless distinguishable and the distinction between them has important implications for the conceptual separation of intelligent and creative responses.

In the simplest terms, the adjectives 'correct' and 'good' apply differentially to the terms 'intelligent' and 'creative'. Intelligent responses are correct; they satisfy objective criteria; they operate within the constraints of logic and reality and thus may be considered right or wrong, true or false. Creative responses, in contrast, are 'good'; they satisfy subjective criteria; although they may not necessarily be limited by the demands of logic and reality, they are responsive, as we shall see, to a wide variety of judgmental standards. Always, though, in the assessment of a response's creativeness or 'goodness' we are aware of the humanness required to make the judgment. As Santayana puts it, 'for the existence of good in any form it is not merely consciousness but emotional consciousness that is needed. Observation will not do, appreciation is required' (Santayana, 1896).

The distinction being made here between the correct and the good is contained, at least implicitly, in many other discussions of the creative process. Guilford, for one, is particularly aware of it when he says, 'There are different bases or criteria by which a product is judged. One is its logical consistency with known facts. Another is its less-than-logical consistency with other experiences' (Guilford, 1957). Guilford goes on, however, to imply that logical criteria are applied primarily to scientific products whereas artistic products are evaluated by less-than-logical standards. In the present distinction no such separation between the evaluation of scientific and artistic products is implied. A piano solo may be as incorrect as the solution to a mathematical equation. A scientific theory may be as good as a great novel. The poet's attempt to write a Spenserian sonnet may be both incorrect and poor, as may the engineer's design for a bridge.

In many of the tests currently proposed to assess creative ability, the failure to differentiate between correct responses and good responses leads to unnecessary confounding. Take as an instance the Remote Association Test developed by Mednick (1963). In this test the *S* is presented with three words, such as 'rat', 'blue', and 'cottage', and is required to supply a fourth word to serve as a kind of associational link between the three stimulus words. For the example given the answer is 'cheese'. Now it is argued that this answer reflects a degree of creativity because it is 'remote' and 'useful', at least in the sense of meeting specified requirements. We would insist that the answer reflects an aspect of intelligence because it is correct. To further illustrate his point, Mednick maintains that the answer 7,363,474 to the question 'How much is 12 and 12?' is original but is not creative because it is not useful. What seems more important, however, is not that the answer lacks usefulness, but that in the absence of a special formal system to justify it, it is wrong. As we would use the term, there are no creative answers to the question, 'How much is 12 and 12?' The constraints of reality implied in the question only allow for correct or incorrect responses.

Mednick introduces the concept of usefulness because he realises, as do most researchers in this field, that the criterion of unusualness or originality is not enough. These qualities, as Guilford's work has so brilliantly demonstrated, do take us beyond the confines of correctness in evaluating cognitive performance. But, as Guilford and others also recognise, the identification of the unusual is, at best, a first step in trying to understand the good.

Creative products and appreciative responses

Frequency and fit

No matter what other positive qualities it might possess, then, we generally insist as a first step that a product be novel before we are willing to call it creative. Indeed, the conjoining of novelty and creativeness is so deeply ingrained in our thinking that the two concepts are sometimes treated as synonymous. The result is that novelty often comes to be used as the most common and, in some of our current paper-and-pencil tests of creativity, the only measure of a product's creativeness.

The application of the criterion of unusualness requires a two-step operation: a comparison of the product in question with other products of the same class and a counting of those comparisons that yield similar or identical products. Further, if such a relative frequency criterion of unusualness is to be applied consistently, some standard must be established to decide how few is few.

The universe of objects with which the judged object is compared also requires definition. The judgment of unusualness is typically made not in terms of *all* other objects of a general class, such as all paintings in existence, but in terms of a greatly restricted subset. Thus, for example, when we say that a child's painting contains an unusual representation of three-dimensional space, our standard of comparison most frequently includes other paintings by children, and we would not make the same judgment if adult work were the standard. When we consider the potential populations against which the unusualness of a particular object might be judged, it becomes apparent that the relative infrequency of the object might alter so dramatically as we shift our frames of reference that the same product could be unique in one population and common in another. Clearly, the choice of an appropiate population or norm group against which to judge a creative work is of utmost importance in applying the standard of unusualness. In short, the infrequency of a response is relative to *norms*, which thus serve as a judgmental standard for evaluating unusualness.

Although the judgment of uniqueness or infrequency may be the logical first step in evaluating the creativeness of a product, if we proceeded no further, the total set of products considered creative by the first criterion alone would be a strange collection, It would include, among other things, all of the products that are simply bizarre or odd. Somehow the mere oddities must be weeded out. This task requires the application of a second criterion, *appropriateness*.

To be appropriate a product must fit its context. It must 'make sense' in light of the demands of the situation and the desires of the producer. Further, when products are complex, their internal elements must also blend together and be appropriate to each other. Thus, both internal and external criteria of appropriateness may be distinguished. Indeed, two separate critical stances arise from the relative emphasis on internal or external views of appropriateness.

The major role of the appropriateness criterion, as it is used here, is to help eliminate from the set of unusual products those that are simply absurd. As a criterion of creativeness, it is used conjointly with unusualness rather than independently, since appropriateness without unusualness would be merely a cliché. At fairly low levels of creative production—as, for example, in paper-and-pencil tests of divergent thinking—the criterion of appropriateness is not too difficult to apply. When a person is asked, for instance, to think of different uses for a common object such as a paper clip and he responds, 'Eat it', the inappropriateness of his response is obvious. When a person is asked to write captions for cartoons and responds by giving the names of colours, again there is little question about the inappropriateness of his reaction. Note, however, that both of these illustrative responses would probably be unusual, if not unique, in a statistical sense. In those cases where the product has to meet some 'reality tests' the application of the appropriateness criterion does not appear to be too difficult.

As products become more complex, however, and more responsive to the needs of the producer than to the demands of the situation, the determination of appropriateness also becomes more complex. There are times, of course, when the judgment is still fairly easy to make. It does not require much aesthetic training to realize that the meter of a poem beginning 'Death is here/Death is there/Death is lurking everywhere. . .' is inappropriate to the theme being treated. Yet at higher levels of creation the entanglement of sense and nonsense can become so involved that the judgment of appropriateness requires a detailed examination by a highly trained person. Even then there is always the chance that the creator, who is often more aware of the demands that gave rise to his product than is the critic, will disagree with the judgment.

Thus, as the term is used in the judgment of creative products, appropriateness deals with much more than the logical fit of the product within its context or the product's elements with each other (although this kind of fit would be the dominant question in many applications of the criterion). There are times, however, when a product violates conventional logic but somehow manages to hang together and to have a logic of its own. Thus, although the judgmental standard for evaluating appropriateness is the *context* of the response, this context must be interpreted psychologically as well as logically and should include the producer's intentions as well as the demands of the situation.

Finally, appropriateness is continuous quality, not a discrete one. It exists in degrees rather than completely or not at all. Considering its lower forms, we recognize the appropriateness of a product that is merely 'responsive'–that bears a clear relation to the environment or to the internal motivations of its producer. Considering its higher forms, we marvel at the way in which the product reflects not only the massive forces that went into its making, but the more subtle influences as well. At low levels of appropriateness we speak of a product as being 'about right', given its sources, the avowed purpose of its maker, and the like. At higher levels we speak of a product as being 'just right'. Indeed, at the very highest level of appropriateness we may experience almost a sense of recognition when we come upon the product. Things are so right they look almost familiar. One critic described such products as having 'the handprint of necessity upon them instead of the quickly tarnished sheen of the merely novel, the fetching precious, the different' (Dickey, 1963). When he comes upon these 'handprints of necessity' the viewer may sometimes feel as if his expectations have been fulfilled, whereas what has really happened is that the product has made him aware of what his expectations should have been.

The transcendence of constraint

Although unusualness and appropriateness may be necessary criteria for limiting the class of potentially creative products, are they sufficient for making required distinctions in quality and level within the class? Among the products that are considered to be both unusual and appropriate some are surely at a higher level of creative excellence than others. One property present in some products but absent, or less obvious, in others is the power to transform the constraints of reality. Some objects combine elements in ways that defy tradition and that yield a new perspective. They literally force us to see reality in a new way. These products involve a *transforma-*

tion of material or ideas to overcome conventional constraints. The question of how much transformation power a product has can serve, then, as a third criterion of creativeness. Just as the unusualness of a product is judged relative to norms and its appropriateness relative to the context, the transformation power of a product would be judged relative to the strength and nature of the *constraints* that were transcended.

The criterion of transformation is more difficult to define than are those of unusualness and appropriateness. At first glance it might seem to be nothing more than an extreme example of unusualness. But it is unusualness with a difference; it is an aggressive unusualness that attacks conventional ways of thinking about things or of viewing objects. In its most dramatic forms a transformation involves a radical shift in approach to a subject or in handling material—the kind of shift, for example, caused by the introduction of the heliocentric theory or Freud's earliest propositions.

The difference between a merely unusual product and a transformation may be approached in another way. Things are often unusual in the purely quantitative sense—in the sense of being the most or the least, the largest or the smallest form of a pre-existing class of objects. (The objects and events described in the *Guinness Book of Superlatives* offer fine examples of this sort of unusualness.) Products that are unusual in this 'record-breaking' sense need not, and usually do not, qualify as transformations. Transformations are not merely improvements, however unusual, on pre-existent forms. Rather, they involve the creation of new forms.

The distinction just made between the concepts of transformation and unusualness requires, now, that a further distinction be made between an object that represents a transformation and one that is merely new. If transformation demanded only a new combination of elements, almost any unique collection of things, however derived, would qualify. Mechanical techniques for supplying these fortuitous combinations could be devised, as in the story of the popular writer who obtained the plots of his novels by spinning a set of wheels on which were written adjectives, nouns, verbs, and the like. Generally, however, these new combinations do not qualify as transformations because they are low in heuristic power. They terminate rather than generate thought. They are the occasion for surprise and laughter, but not for reflection and wonder.

The possibility just alluded to—that the presence of a transformation may be determined in part by its effect on the viewer—raises the interesting question of whether the other two response properties, unusualness and appropriateness, might also have distinguishable effects upon the person who experiences them. Are there, in short, types of aesthetic responses that somehow parallel the criteria of creativeness? This intriguing possibility occured to us at approximately the stage in the development of criteria for creativeness to which the writing has progressed so far. Although recognizing that the job of identifying the criteria was not complete, we decided to delay that task for a time while we attempted to distinguish among possible dimensions of aesthetic response and to relate them conceptually to dimensions of creative performance. In order to give the reader some idea of how the present scheme came into being, we will retrace in the exposition of our ideas the same path that was followed in their development. Thus, for a time, we will put aside the search for more criterion response properties and focus on how the criteria we have already discussed might strike the viewer. The development thus far is summarized in Table 1, which lists the three response properties of creative products and the judgmental standards associated with each.

Table 1

Response properties	*Judgmental standards*
Unusual	Norms
Appropriate	Context
Transformed	Constraints

The impact of the product

Confrontation with an unusual object or event characteristically evokes *surprise* in the viewer. The unusual is attention-getting, it 'catches our eye', its unexpectedness may shock or amaze us. By definition we cannot be prepared for it, except in a very general way. Therefore, the impact of first exposure creates surprize and requires a period of adaptation which the unusual object or event is assimilated into the viewer's experience.

Reaction to the unusual, then, is at its maximum immediately upon exposure and diminishes rapidly thereafter. Though surprize may occur more than once in response to the same object, the second and subsequent exposures never quite match the impact of the first. Objects and events whose value is derived almost solely from their unusualness–such as freaks in side shows or *New Yorker* cartoons–rarely warrant continued viewing.

The quality of appropriateness calls forth in the viewer a reaction, which we will call *satisfaction*, that is akin to the general condition of comfort. This feeling of satisfaction would seem to have two major sources. First, there is the recognition that the demands of the creator, the material, and the milieu have been responded to and that, as was mentioned before, the response is not only right, but is 'just right'. There is thus a recognition of inevitableness about the product, given the context in which it is embedded. Second, there is the recognition that the product is not only 'right' but is complete or sufficient. The first source of satisfaction focuses on the qualitative aspects of appropriateness (i.e. on how *well* the demands are met), the second source of satisfaction focuses on what might be considered the quantitative aspects of appropriateness (i.e. on how *completely* the demands are met). The satisfied viewer's answer to the first criterion is 'just right', his answer to the second is 'enough'.

Products embodying a transformation are likely to be *stimulating* to the viewer. The primary value of such products resides in their power to alter the viewer's conventional way of perceiving or thinking about his world. Whereas confrontation with the unusual requires the viewer to assimilate the product, to make it part of his world, confrontation with a transformation requires that he revise his world. The new element is not only the product itself but the changed environment which has been caused by the product. Metaphorically, a transformation is something like a stone dropped in water. To a person standing on shore the object of interest is not the stone, which quickly disappears from sight, but rather the waves it produces. A transforming object invites the viewer to move out, intellectually, in new directions. A transforming product stimulates the viewer to consider its consequences.

In review, then, the first three criteria of creativeness–unusualness, appropriateness, and transformation–may possibly be the source of three types of aesthetic responses, which may be summarised by the key words: surprise satisfaction, and stimulation. The possibility that different types of aesthetic responses might be isolated gives rise to two important questions which can only be posed here. First, can the aesthetic responses themselves be used to indicate the presence of the qualities that give rise to them? Can we, for example, take the reaction of surprize to be sufficient evidence for the existence of unusualness? Clearly, the answer to this question has extremely important implications for the assessment of creativity. The second question is whether the aesthetic responses are unique to the viewer of the creative product or whether they also appear in the creator himself. Obviously the creator also judges his own product, yet his judgment need not and clearly does not always agree with the verdict of an 'external' judge. Before questions such as these can be examined, however, it is necessary to return to the criteria of creativeness and to ask what else needs to be added to the three that have already been suggested.

The coalescence of meaning

The properties of unusualness, appropriateness, and transformation in varying combinations characterize a large proportion of those things we are willing to call creative. Further, these criteria are applicable to products stemming from very divergent sources, from the scientist and philosopher as well as from the craftsman and artist. Yet despite the power of these criteria there is another important quality that does not seem to be covered by them. This quality, which appears in some of the most highly creative products and which serves as our fourth criterion, deals with the property of *condensation*.

As was mentioned previously, novelty wears out quickly. The history of fad and fashion is, in essence, an elaborate documentation of the short life and relativity of unusualness. In striking contrast, however, to the ephemeral quality of the novel object is the endurance of the greatest creative achievements of man. These we continually seek to re-examine and re-experience. Although the criteria of appropriateness and transformation may partially explain this endurance, something additional seems to be involved.

Products that warrant close and repeated examination are those that do not divulge their total meaning on first viewing. These products offer something new each time we experience them, whether a great work of art or a highly developed scientific theory. They have about them an intensity and a concentration of meaning requiring continued contemplation.

Because confusion and disorder also compel the viewer's attention (as he tries to make sense out of what he sees), it is necessary to distinguish between condensation on the one hand and chaotic complexity on the other. The chief difference would seem to involve the unity and coherence of meaning derived from the condensation as compared with the unrelated and irrelevant meanings derived from disorder. An assortment of debris gathered in a junkyard and the ordered arrangement of the same material by an artist serves to illustrate the distinction being made here. Any meaning derived from the random assortment of junk is fortuitous and is obtained either from a chance association between the elements or from irrelevant associations that the material might stimulate in the viewer. By contrast, the ordered arrangement, if it is worthy of artistic notice, contains more meaning than can be understood at first glance. The colour and shape of the objects, their texture, their spatial location, and their original function all combine to enhance their aesthetic appeal.

In the highest forms of creative condensation the polar concepts of simplicity and complexity are unified. That which at first glance appears simple turns out on closer inspection to possess only *apparent* simplicity. Conversely, that which at first appears to be complex is found to embody a hidden simplicity that binds together the many complex elements. Some of the more successful poems of Robert Frost and some of the paintings of Klee and Miro illustrate well the use of the simple to represent the complex. The reverse situation, that in which initial complexity cloaks a hidden simplicity, occurs frequently in musical works, where repeated listening is often required before the major themes and their variations become apparent.

The condensation achieved by a creative product summarizes essences, and the summary may be expanded and interpreted in a multiplicity of ways–intellectually or affectively, in terms of image or idea. It may be interpreted differently by different viewers or by the same viewer on different occasions. This multiplicity of interpretation and the extensiveness of the expansions generated by the condensation are an indication of its summary power, and an appraisal of *summary power* provides an important judgmental standard for the evaluation of creative condensation.

Parenthetically, it might be noted that the latter two response criteria of transformation and condensation are closely related to the two main mechanisms of the dream work in psychoanalytic theory (Freud, 1955). Thus, the present formulation includes elements that, from a psychoanalytic point of view, illustrate the role of primary process in creative activity (Pine, 1959; Tauber and Green, 1959).

Before leaving this discussion of condensation, two additional points need to be mentioned. The first deals with the increased ambiguity and complexity of the criteria of creativeness as we go from unusualness to condensation, the second deals with their developmental interdependence. At this point the meaning of unusualness appears to be more clearly demarcated than the meaning of condensation, and this is partly because condensation is just a more complex concept than unusualness. Indeed, each of the four criteria, in the order presented here, seems to be more complex than the last. As it is used here, the term complex could be partially translated as meaning 'obscure and complicated', but we wish to stress a more derivative connotation, namely, 'difficult to make judgments of'. We recognize that difficulty in judgment is a consequence of complexity rather than its meaning, but such a distinction is unimportant in the present context. The important point is that judgments of condensation and transformation will be more difficult to make than judgments of unusualness and appropriateness; there will be more differences in viewpoint, and agreement will be more difficult to reach even within schools of thought. This implies that judges will have to know more, or do more, or take more time when judging the criteria of transformation and condensation than they will when applying the other two criteria. In this judgmental sense, then, condensation and transformation are more complex than unusualness and appropriateness.

In addition to this progression in complexity, the four criterion response properties of creativeness are also ordered with respect to developmental interdependence. Each of the four criteria, from unusualness to condensation, is applied in turn conjointly with the previous ones, so that in this sense each is dependent on the ones that precede it. This sequential requirement is only partial, however, in the interdependence of transformation and condensation. The transformations of a creative product must be appropriate and unusual; the condensation must also be appropriate and unusual, but it may not always represent a transformation. If we accept unusualness and appropriateness as necessary properties for a product to be considered creative, then the hierarchical ordering of transformation and condensation provides an additional basis, along with degrees of variation within each of the response dimensions, for distinguishing levels of creative attainment within the class of creative products.

Having added a fourth response property, condensation, we return to a consideration of aesthetic responses to examine the possible effect of this new criterion upon the viewer. So far our hypothetical viewer has been surprised by unusualness, satisfied by appropriateness, and stimulated by transforming qualities of creative products. The question now is how will he react when he confronts a product of high condensation, one that exists on many levels of meaning. The definition of the criterion itself almost contains a description of the experiencing viewer. A condensed product is an object worthy of pondering; it should be examined slowly, carefully, and repeatedly. In a word, the viewer is called upon to *savour* a condensation. The surprize, the satisfaction, and the stimulation that characterize responses to other criteria of creativeness are present as well in the response to a condensation, but there is an important difference. In the reaction to a condensation these other responses are enduring and somewhat intensified. Surprise occurs not only on the first encounter, but also on subsequent ones as new and unusual aspects of the product are discovered; satisfaction deepens with repeated exposure as the appropriateness of each element in the product is more fully revealed; stimulation is enriched as each new reaction to the product builds on those that have preceded it. It is this continued freshness of the product and of the viewer's response to it that makes it an object worthy of savouring.

With the addition of condensation to the list of response properties and of savouring to the list of aesthetic reactions, we complete our discussion of creative products and appreciative responses (see Table 2 for an outline of the development thus far). One further point needs to be made explicit, however: just as the criterion response properties are relative to judgmental standards, so are the aesthetic responses.

Table 2

Response properties	*Judgmental standards*	*Aesthetic responses*
Unusualness	Norms	Surprize
Appropriateness	Context	Satisfaction
Transformation	Constraints	Stimulation
Condensation	Summary power	Savouring

The degree and character of surprise is a function of the norms of expectation in much the same manner as is unusualness. Similarly, satisfaction is relative to the context, stimulation to the nature and strength of the constraints, and savouring to summary power.

Personal qualities and cognitive styles

Among psychologists, creativity has frequently been thought of as a single dimension or at least as a unified cluster of traits, resembling–and to some extent overlapping with–general intellectual ability. Some researchers, particularly factor analysts such as Guilford, have isolated a set of cognitive factors related to different aspects of intellectual production, but the relation of these factors to creative production is less well delineated. Other investigators have argued that creative behaviour depends as much on personality as on cognitive power, and they have accrued evidence to show that the highly creative person is more impulsive, makes greater use of fantasy processes, is more tolerant of ambiguity, and so forth, than is his less creative peer.

At the level of everyday experience it is recognized that creative expression of the highest quality tends to come from people who limit their efforts to a single mode of expression. The great scientist is generally not also the great poet, the great painter is not also the great dancer, and even the great composer is not also the great conductor. Moreover, in the eye of the public, personality characteristics tend to be associated with a *medium* of expression rather than with the level of expression or any other features of creative production. Thus, in terms of personality characteristics, the outstanding poet looks more like the incompetent poet than he does like the outstanding scientist, who, in turn, looks more like his less competent colleagues than he does like the superb composer. Both the professional view and the public view, therefore, acknowledge that the maker of creative products looks different from his fellow men, but there is little agreement on the form this difference takes. Is the difference chiefly cognitive or does it involve personality components as well? Is the difference linked to a mode of expression entailing, as it were, a way of life, or is it more closely related to stylistic qualities of the products themselves? Each view of the creative process implies a position about the nature of the creator, which in turn helps to determine the types of variables emphasised in the assessment of creativity.

Within the present context there is a set of personal qualities that may be considered to match, so to speak, the response properties used as criteria of creativeness. The present view is not meant to imply that trait names should be used as labels in characterising people. An answer to that question depends on such matters as the dominance of the trait in question, its distribution in the population, and so on. Rather, we wish to suggest a relation between personal qualities and properties of the creative response.

The question of whether personal qualities are necessary for the production is still another matter. Do personal qualities 'cause' the appearance of the creative criteria? Might an infrequent or appropriate response occur by chance or must it always occur as the result of a particular human condition? Again, the view taken here does not prejudge the answer to this question, although it does imply that the consistent production of creative responses cannot occur by chance and, further, that the 'cause' of such a phenomenon entails psychological as well as social and environmental influences.

In the present view the person who consistently produces infrequent or unusual responses is thought of as being highly *original*. The relativity of the judgments of unusualness would also affect the judgment of personal originality. Persons whose responses look quite unusual when judged against one standard and who would, therefore, be thought of as original, might lose the halo of originality when the judgmental norms are changed. The personal quality of originality is an attainment, in the sense that it is inferred from the repeated production of unusual responses. In addition, there are several personal qualities that are predisposing to originality, that increase the likelihood of unusual responses but do not guarantee them. These predisposing characteristics include intellectual abilities, cognitive styles, motives, and values. For example, ideational fluency, impulse expression, and cognitive styles of tolerance of unreality, tolerance of inconsistency, and the like would appear to be likely candidates for predisposing the individual towards originality (Barron, 1955).

The production of an appropriate response would seem to be accomplished most easily by a person who is highly *sensitive* to the demands of his environment and to the subtleties of the material with which he is working. His sensitivity may result from a conscious *analysis* of the relations between elements, but it need not. There is some evidence that at least in some fields the most sensitive people cannot articulate their awareness with any degree of precision. Their sensitivity is *intuitive*; the person who behaves intuitively is sensitive to cues that cannot be identified verbally. Thus, both an analytic attitude and an intuitive attitude could serve as predisposing cognitive styles for sensitivity.

Transformation involves the transcending of traditional boundaries and limitations. Personal qualities that would seem to contribute to the production of transformations are of two sorts, cognitive and non-cognitive. Relevant cognitive qualities are those that deal with the stability and the fluidity of conceptual systems, intellectual categories, and the like. The production of transformations reflects *flexibility* and calls for qualities not unlike those involved in breaking a 'set'–as in the Luchins water-jar experiment. In perceptual terms this type of intellectual fluidity would be reflected in ability to perceive objects in their own right–independent of their symbolic representation, their stereotyped function, or their relatedness to the immediate needs of the viewer. This kind of perceptual freedom has been given the label 'allocentric perception' by Schachtel (1959). At the cognitive and ideological level, this flexibility has been called 'openmindedness' by Rokeach (1960).

The non-cognitive qualities that contribute to the production of a transformation include a playful attitude towards reality and a willingness to expose (even to flaunt) ideas, attitudes, and objects that violate tradition. An attitude of playfulness, a desire to toy with reality is important because it would appear that most transformations come as a discovery on the heels of many trials. Thus to some extent transformations involve an element of 'luck', but it is the kind of luck that cannot occur without the predisposing attitudes that lead to experimentation and other forms of intellectual play. No single adjective adequately summarizes the personal qualities that contribute to the production of transformations. The word that comes closest to describing the cognitive qualities and one that at least does not do violence to the non-cognitive qualities is the adjective *flexible*.

Just as a condensation may often contain a paradoxical union of simplicity and complexity, so does the production of the condensation call for fusing of contradictory personal qualities. There is first the coalescence of personal and universal concerns with strangely paradoxical result of losing in self-awareness while gaining in total awareness. Although the product may bear the marks of its maker, his personal needs and interests have been consumed, as it were, by the more abstract relevance of the product itself. Second, there is the intimate interplay, even fusion, of thought and feeling that contributes to a condensation. Even the most 'logical' production–such as an elegant mathematical solution to a problem–demands an openness to thoughts and feelings that are only dimly perceived and that may be incapable of precise logical statement. Third, there is an alternate blending of working style, a cyclic pattern of patience and passion, of reflection and spontaneity, a continual shifting from total acceptance of one's ideas and actions to a critical rejection of them. In a very useful discussion of the creative process, Erich Fromm (1963) distinguishes between two phases of productive behaviour. According to Fromm, the first phase is essentially feminine in quality. It is the birth-giving phase, the moment of conception. Following this, a more masculine phase occurs during which the creator must hone and polish his work to ready it for social judgment. In this second phase irrelevances and superfluities are eliminated, the uniqueness of the work is more sharply defined, and its content is more effectively expressed. The central theme underlying the production of a condensation involves, then, the unification of things that are normally thought of as separate or at least distinguishable–producer and product, personal and universal concerns, cognitive and affective processes, masculine and feminine styles of work.

The complexity of condensations and the personal qualities from which they emerge do not seem to permit the use of a single adjective to describe the person who is most apt to produce condensations. Yet, in the interests of symmetry and as a mnemonic device for summarizing the conceptual scheme presented here, we searched for a word that would come close to capturing both the substance and the sense of our description. The word that in our opinion does the most justice to what has been said about the producer of condensations is *poetic*. Webster's *New World Dictionary* describes a poem, and thus indirectly a poet, as 'expressing facts, ideas, or emotions in a style more concentrated, imaginative, and powerful than that of ordinary speech'. The terms 'concentrated', 'imaginative', and 'powerful' are relevant here and apply of course to many products that would not be commonly thought of as poems. It is this more global referent of poetic qualities that we imply when we characterize people who create effective condensations as poetic. The categories are summarized in Table 3.

Table 3

Predisposing cognitive styles	*Personal qualities*	*Response properties*	*Judgmental standards*	*Aesthetic responses*
Tolerances of incongruity, of inconsistency, etc.	Original	Unusualness	Norms	Surprise
Analytic and intuitive	Sensitive	Appropriateness	Context	Satisfaction
Openminded	Flexible	Transformation	Constraints	Stimulation
Reflective and spontaneous	Poetic	Condensation	Summary power	Savouring

The creative person, his product, and the world's response to it combine to form the drama of human invention. The transaction among these three elements is intimate and our understanding of any one enhances our understanding of the other two. However, though at present far from it, social science must work towards a conceptualization that serves to unify the psychological, aesthetic and social aspects of this phenomenon. It is not enough, of course, to offer pious hopes for a future ecumenical council whose job it will be to unify these divergent views of the creative process. Two kinds of preparatory work are necessary before such a grand design is sought. First, we must add greater detail to early attempts at unification–such as the one presented here. The present scheme, at best, is an incomplete outline that must be elaborated more fully if its potential contribution is to be realized. Second, we must initiate empirical studies that will give weight to these efforts at theory, thus preventing them from deteriorating into mere word games.

In the final analysis, however, it is well to remember that theories of creativity are themselves creative products. As such, they must abide by the same laws as those they are designed to unearth. A realization of this fact should temper our zeal in advocating any single prescription for how best to proceed. The day on which we are certain about how to construct a theory of creativity will also be the day on which we are certain about how to construct a poem.

References

BARRON, F. (1955) 'The disposition toward originality', *J. abnorm. soc. Psychol., 51,* 478-85.

DICKEY, J. (1963) *The New York Times Book Review,* 22 December, 1963, p. 4.

FREUD, S. (1955) *The Interpretation of Dreams.* New York: Basic Books.

FROMM. E. (1963) 'The creative attitude', in H. ANDERSON (Ed.) *Creativity and Its Cultivation.* Michigan State Univ. Press.

GOLANN, S.E. (1963) 'Psychological study of creativity', *Psychol. Bull., 60,* 548-65.

GUILFORD, J.P. (1957) 'Creative abilities in the arts', *Psychol. Rev., 64,* 110-18.

MEDNICK, S.A. (1963) 'The associative basis of the creative process', in M.T. MEDNICK and S.A. MEDNICK (Eds.) *Research in Personality.* New York: Holt, Rinehart & Winston.

PINE, F. (1959) 'Thematic drive content and creativity', *J. Pers., 27,*136-51.

ROKEACH, M. (1960) *The Open and Closed Mind.* New York: Basic Books.

SANTAYANA, G. (1896) *The Sense of Beauty.* New York: Scribner's.

SCHACHTEL, E.G. (1959) *Metamorphosis.* New York: Basic Books.

TAUBER, E.S. and GREEN, M. R. (1959) *Prelogical Experience.* New York: Basic Books.

27 Creativity and curriculum structure

E. Ogilvie

Introduction

A consideration of the literature devoted to discussions of the relationship between creativity and educational environment reveals a situation which is in need of much closer examination than it has received hitherto. In the first place, there is general presumption that 'informal' or 'progressive' schools provide an atmosphere conducive to the development of creative potential, and that the 'formal' or 'traditional' environment is highly inimical to the creative child (Torrance, 1962). It is difficult to ascertain how far terms like 'progressive', 'traditional' and 'creative' indicate relatively simple and generally aççepted concepts, but we refer here to those schools which emphasize so-called creative activities,self-initiated learning and discovery on the one hand, and to schools which concentrate on the didactic and authoritarian approach on the other (see Barker Lunn, 1970). The attributes of the creative child are listed by Torrance (1962) and described by traits in Guilford (1959). Basically the creative child is a nonconformist who questions authority and may even hold conformists in some disdain. He is an individualist with a proclivity for experimentation.

There are no agreed theoretical bases to support the common view that traditional environments are detrimental to creativity. It may be that its origins lie in accounts of the unhappy childhood experiences of a number of highly creative individuals (Goertzel and Goertzel, 1965; Burt, 1970). Some firm evidence is provided by Haddon and Lytton (1968) who found eleven-year-old English children to be significantly superior in divergent thinking abilities when taught in 'informal' schools. These authors cite other sources, Torrance (1962), Vernon (1964), Spears and Hilgard (1964), but point out that there is really very little evidence which directly tests and supports the hypothesis that 'informal' schools provide the best climate for creativity. Guilford (1967) offers something of a theoretical background supporting the general viewpoint but his remarks are based largely on generalities. Stoddard (1959) states the dogma lucidly but again without a single reference to hard evidence.

Torrance's view is clearly that an informal atmosphere fosters creativity since, on the basis of wide experience of creativity testing, he cites approvingly (1964) the comment 'rigid authority deprecates ideas offered by pupils'. It is interesting, therefore, to notice that Torrance also believes the relationship between intelligence and creativity depends to no small

---

Source: *Educational Research,* 1974, *16,* 126-32.

extent upon the degree of structuring displayed by the school curriculum. Thus he seeks to explain Hasan and Butcher's (1966) failure to differentiate intelligence from creativity by arguing that the Scottish secondary school environment is too formal to permit the true facts to reveal themselves and that it brings about an artificially high correlation between creativity and intelligence quotients.

Dacey, Madaus and Allen (1969) however report only very low IQ/CQ correlations in their samples of Irish children, and it seems unlikely that the Irish secondary grammar schools were significantly less traditional than their counterparts in Scotland.

Nevertheless, these authors by way of explanation for their low intercorrelations also appeal to factors linked to curriculum structure, and suggest that higher creativity-intelligence intercorrelation may be the result of 'an interaction between method factors relating to both measures rather than to a high relationship between the two traits'.

Haddon and Lytton (op. cit.) specifically predict higher correlations between creativity and intelligence in the informal schools; a view diametrically opposed to that given by Torrance. Haddon states that in the informal schools, 'because of their particular approach to learning, the link between the two aspects would be maintained to a greater extent' and continues, 'the idea is more forcibly expressed by saying that formal education will tend to destroy this connection by putting a premium on convergent thinking and conformist behaviour'.

It is noteworthy that none of the authors cited appears to be very clear about the nature of the 'link' between intelligence and creativity. Indeed the explanation offered amounts to little more than a restatement of the facts as they appear in the particular research findings; and there can be little doubt that the theoretical interpretation of interrelationships of creativity, intelligence and curriculum structure is unsatisfactory.

The present situation thus presents certain difficulties with regard to experimental exploration in the field. It is usual to match schools on some variable or variables before comparing their performances on creativity tests. Matching is necessary when (*a*) the matching variable correlates significantly with the criterion variable, (*b*) the criterion variable is in some real sense dependent upon the capability measured by the matching variable and (*c*) the independent variable and the matching variable are unrelated. Intelligence is frequently seen as meeting these conditions, and matching groups on verbal reasoning tests has become almost standard procedure in connection with research on teaching methods and related problems.

The works of Vernon (1964) and Pidgeon (1970) demonstrate that IQ is no longer thought of as an unchanging, innate attribute. Many of the experiments they cite show clearly that there is little warrant for treating intelligence as the variable independent of curriculum structure. As Dacey *et al.* (op.cit.) suggested, both intelligence and creativity can be considered as variables which are likely to be appreciably dependent upon the teaching methods employed in schools. Moreover, it is not unreasonable to guess that very informal schools are unlikely to provide environments conducive to the production of high Verbal Reasoning Quotients whereas formal schools may frequently approach optimum conditions in this regard (Wiseman, 1966; Butcher, 1968). If this latter hypothesis is true; and there is evidence to suggest that it is (Surrey Education Committee, 1968), and, in addition, a positive correlation is found between intelligence and creativity, then it appears that to match formal and informal schools on verbal reasoning is effectively to favour the informal school in any comparisons on creativity. Where the informal environment favours creativity the apparent effect will be exaggerated, and where the formal school favours creativity the effect will tend to be masked. Haddon and Lytton take no account of this possibility. Indeed a special point is made of the fact that informal schools have a mean VRQ slightly, but not significantly, below that of the formal schools, yet, in spite of this, the informal schools achieve results which are significantly superior'.

The conclusion we draw so far is that, in the absence of other information, there is no point in matching schools on VR tests in research into the effects of curriculum structure and that this procedure might be particularly misleading where a very wide range of environments is deliberately built into the research design, a problem discussed by Barker Lunn (1970).

This point introduces a second consideration which deserves attention. Researchers often content themselves with an exploration of the informal–formal dimension in terms only of a dichotomy. This is consistent with the form in which the problem is usually posed, and may be justified on the grounds of simplification, larger numbers in the groups, and so on. It can be contended, however, that such simplification may bring special problems of its own. For example, if it should turn out in research of this kind that informal schools appear to be more conducive to creativity than formal schools then teachers might readily infer a maxim such as 'the more informal the better'. This would be unfortunate since it could be maintained that some intermediate point of the formal/informal continuum might be best and that the postulated linear relation between informality and creativity was only true for part of what was in fact a curvilinear relationship.

The difficulty is perhaps best revealed by means of a figure (see Figure 1).

In the following sections a theory of creativity is outlined and relevant experimental findings are considered in the light of the foregoing discussions.

A theory of creativity

Research into creativity and curriculum structure must be based upon a

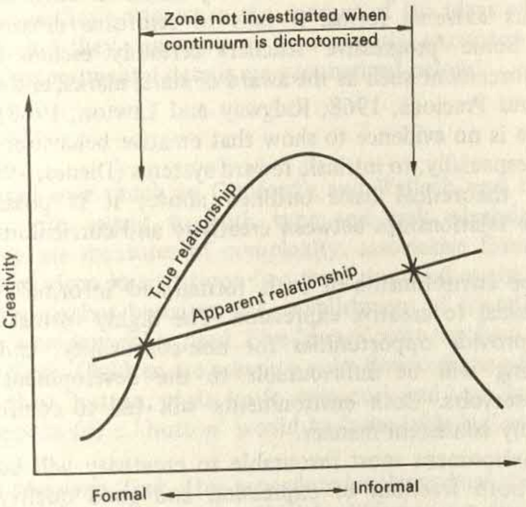


Fig. 1 Creativity and informality; a curvilinear relationship?

theory of creativity which is made quite explicit, and the theoretical standpoint adopted in the present research derives from the Associationist Theory of Creativity as outlined by Mednick (1962). We postulate that creativity depends firstly upon the size and quality of the association reservoir. Mednick argues that an individual without the requisite elements in his response repertoire can hardly be expected to combine them in a creative solution: a child who has not learned his numbers will never combine them in novel number series. He further suggests that the greater the number of associations which a person has constructed amongst the requisite elements, the greater the probability of discovering new combinations.

Creativity depends secondly upon the relationship between the creative personality and the curriculum structure which helps to shape that personality. The general view of the effect of school environment on creativity noted earlier takes account only of this second point since it stresses simply the deleterious and inhibitive effects of authoritarianism and conformity pressures. It neglects or ignores the fact that a very 'progressive' or unstructured school situation might fail to teach children those very elements of learning and knowledge which are essential to the solution of problems included in creativity tasks.

It can reasonably be argued therefore that, beyond a certain point, a curriculum becomes so unstructured that it fails to provide a vital condition for creativity beyond mere freedom, namely, that of a basic achievement level or association reservoir. Moreover, it seems highly probable that some creative children are inhibited in the very informal environment simply because, in its extreme forms, it fails to reinforce creative behaviour consistently. Some 'progressive' teachers certainly eschew any form of extrinsic reinforcement such as the award of stars, marks, and verbal praises (see Brown and Precious, 1968; Ridgway and Lawton, 1968), despite the fact that there is no evidence to show that creative behaviour is responsive only, or even especially, to intrinsic reward systems (Dienes, 1964).

Given the theoretical basis outlined above, it is possible now to summarize the relationships between creativity and curriculum structure as follows:

(1) Extreme environments of both 'formal and 'informal' types will be relatively inimical to creative expression. The highly formal environment will fail to provide opportunities for non-conformity, and the highly informal setting will be unfavourable to the development of sizeable association reservoirs. Both environments will fail to reinforce creative behaviour in any consistent manner.

(2) The environment most favourable to creativity will be that which provides for both freedom of expression and good quality association reservoirs. It will be especially favourable if extrinsic rewards are attached to creative behaviours.

(3) An environment so lacking in structure as to approach a state of anomie will be worst of all since it provides neither freedom nor basic achievement skills. It quite fails moreover to provide any reinforcement outside a random schedule.

(4) Only very extreme environments are likely to be totally hostile to creative output since some creativity tasks require very little in the way of school achievement and verbal learning. At more intermediate levels formal environments might provide relatively congenial conditions for success on tests requiring relatively high levels in basic skills. Informal curricula on the other hand might be relatively more effective in areas where the basic achievement levels required are relatively low. It is convenient at this point to note that in any thoroughgoing investigation it would be important to examine the exact nature of the differences between the various intermediate grades of school climate in terms of the individual creativity tasks and the minimum levels of achievement necessary to their performance.

The Associationist Theory predicts a positive correlation between intelligence and creativity, regardless of curriculum structure and supports a 'threshold' theory with regard to this interrelationship. It suggests that the influence of structure on CQ/IQ intercorrelations is likely to be highly complex, and warns against easy generalization in this regard.

Details of the investigation

The investigation below sheds some light on these theoretical problems though it was not designed specifically to answer them. The findings are therefore not without interest in the context of the ideas which have been put forward, and they constitute as least a partial excursion into the field. The essential experimental details are summarized below:

Creativity tasks

The tests used owe much to Guilford's and Wallach and Kogan's (1965) conceptions with regard to both type and task administration. Taken together they are measures of complexity, associative fluency and originality and were given in situations free from time and examination stresses. Game-like approaches based on the establishment of friendly rapport with the children were invariably used. Five 'games' were involved.

1. *The Uses Task.* Children are asked to write down all the possible uses for a newspaper, shoe, button, chair, knife, tyre, cork and key. An example of an unusual response for a 'button' would be 'raise table leg' as compared with 'fasten coat'.
2. *The Line Meanings Task.* This procedure involves visual stimulus materials whereby the child is confronted with eight different line diagrams and asked to generate meanings relevant to the line in question.
3. *The Number Series Task.* This instrument is a form of the familiar Make Up Problems Tests (Getzels and Jackson, 1963) and requires the child to construct a set of different mathematical series. It is similar to Westcott's (1968) assignments and is scored for both fluency and orignality.

The contrast between commonplace and original responses can be seen in the examples given below:

(i) 2, 4, 6, 8, 10. . .
(ii) 15, 13, 11, 9, 7. . .
(iii) 6, 7, 12, 13, 24, 25, 48. . .
(iv) 50, 47, 44, 42, 40, 43, 39, 36, 33, 31, 29

4. *The Symbol Series Task.* This instrument is similar to Task 3, but symbols and/or letters are used instead of numbers. It is scored for fluency only and an example would be: □ ○ △ □ ○ △ □.
5. *The Kites-Fishes Puzzle Design Task.* Children are here required to construct a puzzle commonly found in children's puzzle books. Three children are envisaged as flying three kites or fishing for three fishes and the child must draw in the three tangled strings so as to create the most complex puzzle he can. He is further required to design original kites and fishes so that complexity and originality scores can be constructed.

The schools

Five schools well known to the author, several colleagues and an LEA administrator were selected and ordered on the 'structural' dimension as follows:

School	*Type of environment*
A	Highly structured (Formal)
B	Midroad I
C	Midroad II
D	Unstructured (Informal)
E	Indeterminate (Anomie)

The main criteria used in identifying the 'structure' dimensions are indicated in Table 3.I of Barker Lunn (op. cit, p. 27).

School E is described as in a state of anomie since the general organization lacked coherence and it presented features of both types of environment always attenuated by an air of casualness.

The effects of differential home backgrounds are confounded in the findings since school E contained wholly working-class children. The other schools were more comparably homogeneous in terms of parental social status.

The subjects

The children comprised the fourth-year groups (ten- to eleven-year-olds) in each school. The numbers varied somewhat between the tasks owing to the fact that testing sessions had to be spread over the whole summer term. The school B group was tested one year later than the other groups for practical reasons irrelevant here. The total number involved ranged from 131 to 161 children.

Results and discussions

The results are set out in the following tables.

The strength of the highly significant relationship amongst creativity scores (Table 1) is indicated by Haggard's (1958) R=0.20 or the correlation ratio E=0.44. These results are thus consistent with the main hypotheses put forward earlier in that neither of the less structured environments produce children with higher scores on the creativity tests and the most uncertain classroom organization gives rise to the lowest scores.

The question as to how far these findings conflict with those of other workers is difficult. The fact that our extremely undirected school E seems least conducive to the creative development of children can readily be reconciled with the relative success of Haddon and Lytton's informal group (1968) by postulating that our midroad environments are akin to their informal type. Their school D seems in fact to resemble our school C quite well, and the degree to which our school D is fairly extreme may be judged from the headteacher's remark, 'We teach our pupils not to be able to do

Table 1 *Comparison by schools of the mean creativity aggregate scores: summary data and analysis of variance*

Schools	A	B	C	D	E
	Structured	*Midroad I*	*Midroad II*	*Unstructured*	*Anomie*
Means	206	231	227	199	171
Source	s.s.	df	ms	F ratio	
Between	53,292	4	13,323	7.54 (0.01)	
Within	222,702	126	1,767		
Total	275,994	130			

intelligence tests'. This explanation of the discrepancy between present findings and those of Haddon and Lytton is attractive since alternative proposals such as incomparability of tests or testing situations do not seem to be particularly impressive. It is important therefore to stress that acceptance of the view put forward here implies that qualifications to Haddon and Lytton's major thesis are very important. The proposal that 'informal progressive teaching would promote these (divergent thinking) abilities more than formal subject-centred teaching' can at best be only partially true, and could easily be a dangerous half-truth at that. It has been made clear here that the 'structure' dimension involved both requires and is susceptible of a greater measure of subdivision than a simple dichotomy provides. The main reason for greater subdivision is that the relation of creativity to curriculum structure is in all probability much more complex than it appears and oversimplification and generalization can only do more harm than good. This is particularly true in an area where many educators appear to have already formed opinions in the absence of empirical evidence and seem willing therefore to fasten on any research findings which appear to support their views (Hudson, 1970).

The evidence of the present research to some extent goes against popular opinion. It shows the formal schools tested were less inimical to creative potential and expression than were the types of progressive schools tested, and that this may be more true of some curriculum areas than others. There can be little doubt that the Midroad environments were best in general terms, but we suggested earlier that this might be due merely to their failure to suppress rather than to their making any really positive contribution to the development of creative potential. Further research in this area ought to specify more closely than has been possible here, the exact nature and range of the environments involved, using perhaps the checklist given by Barker Lunn as a basis for more objective categorizations. It ought to investigate individual differences in response to differing environments in order to fill in some of the theoretical gaps to which attention has been drawn.

The evidence with regard to the interrelationships of intelligence and

Table 2 *Intercorrelation of VHQ and CQ in terms of curriculum structure*

Curriculum Type	*Structured*	*Midroad I*	*Midroad II*	*Unstructured*	*Anomie*	*Total Group*
Correlation rhc	0.49**	0.17	0.41*	0.31	0.17	0.54**
N	31	41	20	20	19	131
Mean CQ	101.1	115.1	109.5	99.6	96.2	106.1

**0.01 *0.05

creativity as reported in Table 2 suggests that different types of children are differentially creative in different environments.

The evidence unfortunately is meagre, even confusing, and the main value of the table is that it underlines the need for further research oriented towards solving the theoretical issues raised earlier. Depending upon which group is chosen, either Torrance's or Haddon's position receives some support; and yet both are contradicted. It seems that in all probability the interrelationship between VRQ and CQ depends primarily upon the general intelligence level of the group concerned and only to a relatively slight extent on curriculum structure as such. There is now a growing body of research which amply supports the 'threshold' concept of this relationship, (Haddon and Lytton, op.cit.; Getzels and Jackson, op.cit.; Hudson, 1966) and Wallach and Kogan's claim to have discovered an independent dimension can be discounted on the grounds that their group averaged *c.* 120 IQ. The present results given in Table 3 display the 'threshold' phenomenon quite nicely and are thus in line with the 'Associationist' Theory of Creativity. They suggest that the 'threshold' is, however, well below the 120 IQ level often proposed (McKinnon, 1962).

Table 3 *Intercorrelations of creativity and intelligence*

	Range	*VRQ Mean*	*SD*	*Correlation CQ VRQ*	*N*
Creativity:	Total	105.8	15.2	0.547	131
	Upper Half	112.8	12.2	0.133	66
	Lower Half	99.0	14.7	0.537	65
Intelligence:	70-139	105.85	15.18	0.547	131
	107-139	117.99	8.40	0.108	66
	70-107	93.90	10.00	0.484	65

Conclusion

The relationship of creativity to curriculum structure is more complex than has been admitted in previous work. The commonly held opinion that 'informal' environments are generally more congenial to creative expression requires careful qualification and in our present state of knowledge any generalization is dangerous. It is suggested that until individual differences in creative response to various degrees of curriculum structure are more fully investigated the most conducive environment is that characterized here as 'midroad'.

The 'threshold' theory of the interrelationship between intelligence and creativity receives considerable support from present findings which are wholly in accord with those of Haddon and Lytton at this point. The threshold is likely, however, to be well below the 120 IQ level sometimes suggested. The Associationist Theory carries no clear implications for the interrelationships of creativity, intelligence and curriculum structure and is in need of extension. Meanwhile, there is little warrant for attempting to explain conflicting findings with regard to creativity-intelligence correlations in terms of differences in curriculum structure. A far more plausible explanation would seem to derive from the concept of an intelligence threshold.

References

BARKER LUNN, J.C. (1970) *Streaming in the Primary School.* Slough: NFER.

BROWN, M. and PRECIOUS, N. (1968) *The Integrated Day in the Primary School.* London: Ward Lock Educational.

BURT, Sir C. (1970) foreword to KELMER PRINGLE, M.L. *Able Misfits.* London: Longmans.

BUTCHER, H.J. (1968) *Human Intelligence.* London: Methuen.

DACEY, J., MAUDUS, G. and ALLEN A. (1969) 'The relationship of creativity and intelligence in Irish adolescents'. *Br. J. Educ. Psychol., 39,* 261-7.

DIENES, Z.P. (1964) *Mathematics in the Primary School.* London: Macmillan.

GETZELS, J.W. and JACKSON, P.W. (1963) *Creativity and Intelligence.* New York: Wiley.

GOERTZEL, M. and GOERTZEL, V. (1965) *Cradles of Eminence.* London: Constable.

GUILFORD, J.P. (1959) 'Traits of creativity', in *ANDERSON, H.* (Ed.) *Creativity and its Cultivation.* New York: Harper & Row.

GUILFORD, J.P. (1967) *The Nature of Human intelligence.* London: McGraw-Hill.

HADDON, P.A. and LYTTON,, H. (1968) 'Teaching approach and the development of divergent thinking abilities in primary schools', *Br. J. educ. Psychol., 38,* 171-80.

HAGGARD, E.A. (1958) *Interclass Correlation and the Analysis of Variance.* New York: Holt, Rinehart & Winston.

HASAN, P. and BUTCHER, H.J. (1969) 'A partial replication of Getzels and Jackson's study',*Br. J. Psychol., 57,* 129-35

HILDRETH, G. (1966) *Introduction to the Gifted.* New York: McGraw-Hill.

HUDSON, L. (1966) *Contrary Imaginations.* London: Methuen.

HUDSON, L. (1970) As reported in *Times Educ. Suppl.*, 24.7.70, p.9.

MEDNICK, S.A. (1962) 'The associative basis of the creative process', *Psychol. Rev., 69* 3.

MCKINNON, D.W. (1962) 'Nature and nurture of creative talent', *Am. Psychol., 17,* 487-8.

PIDGEON, D.A. (1970) *Expectation and Pupil Performance.* Slough: *NFER.*

RIDGWAY, L. and LAWTON, I. (1968) *Family Grouping in the Primary School.* London: Ward Lock Educational.

SPEARS, P.S. and HILGARD, E.R. (1964) 'The teacher's role in the motivation of the learner', in *Theories of Learning,* 63rd NSSE Yearbook, Univ. of Chicago.

STODDARD, G.D. (1959) 'Creativity in education', in ANDERSON, H. (Ed) *Creativity and its Cultivation.* New York: Harper & Row.

SURREY EDUCATION COMMITTEE (1968) Report of the Research Association, *To Stream or not to Stream.*

TORRANCE, E.P. (1962) *Guiding Creative Talent.* Englewood Cliffs. New Jersey: Prentice Hall.

TORRANCE, E.P. (1964) *Role of Evaluation in Creative Thinking.* University of Minnesota.

VERNON, P.E. (1964) *'Creativity and intelligence', Educ. Res., 6,* 163-9.

WALLACH, L.M. and KOGAN, H. (1965) *Modes of Thinking in Young Children.* New York: Holt, Rinehart & Winston.

WESTCOTT, M.R. (1968) *The Psychology of Intuition.* New York: Holt, Rinehart & Winston.

WISEMAN, S. (1966) 'Environmental and innate factors and educational attainment', in MEAD, J.E. and PARKES, A.S. (Eds.) *Genetic and Environmental Factors in Human Ability,* 65-80. Edinburgh: Oliver & Boyd.

Index

abilities or intelligence
  and achievement 78, 81, 84, 107–8, 300, 346–7, 351
  and creativity 336–40, 368–70, 385–6, 388, 391–3
  and effectiveness of preschool intervention 174, 175
  and fear 123, 124
  and field dependence/independence 293
  and interests 84
  and preschool education 174, 177, 185, 187
  and race 345–56, 361–3, 366
  and relationship of personality to achievement 78, 80, 81, 84, 107–8
  and relationship of social class to aspirations 146
  and test situation factors 352
  measurement 333–4
  of managers 355
  of parents and children 359, 361, 365
  of rural *v.* urban people 366–7
  regression effects 358–9
  theories 294, 300, 307–9, 329–32, 341, 342, 363–4, 365, 369
academic discipline and relationship of personality to achievement 101–2
achievement (including performance and attainment)
  and child-rearing practices 143–4
  and creativity 338–9, 340
  and intelligence or abilities 78, 81, 84, 107–8, 300, 346–7, 351
  and parental encouragement 145–6
  and personality 78–85, 96–110
  and race 346–7, 351, 353
  and social class value systems 143, 146–55
  measurement methods 105–6
  of high IQ and high creative children 338–9
  on inductive reasoning problems 317–19
achievement of competence as an intrinsic learning motive 134–7
achievement motivation 150
achievement want and social class 151
Adelson, J. 151
adjustment-anxiety factor of personality 90–1
adolescent development 191–203
adopted or fostered children's IQ
  and parents' IQ 362, 365
adult differences in psychological development
  and adolescent maturity status 201
aesthetic responses to creative products 373–5, 377–8
  as indicators of creativity 375
affective mode of defining equivalence 206, 207
affiliation needs and maturity status 198–9
age and cognitive functioning 302–3, 363–7
  and correspondence 221, 222–36
  and equivalence formation 204–19
  and fear 123
  and relationship of personality to achievement 80, 101. 106, 108–9
  and understanding of invariance 242–53
aggression and maturation status 193, 197–8
aggressiveness repression 41

aims in reading research papers 14–16
Allen, A. 385
amputees castration anxiety 69–70
anal phase of sexual development 43
analysis of variance 11, 81, 175, 243, 245–6, 250
analytical IQ
  and field dependence/independence 294, 300, 307
  of retarded children 308
Anthony, A. 69
anthropological approach to Oedipus complex 60
anticathexes or resistances 50, 51
anxiety or conditioned drive 117–18, 123
appropriateness to context of product
  and aesthetic response of satisfaction 374, 375, 378
  and sensitivity as property of creative individuals 380
  as judgmental standard for evaluating creativity 371–2
arithmetic teaching in the lower grades 158–60
  inefficiency of schools 160
aspirations (including ambitions) and social class value systems 146–8, 151–3
Associationist Theory of Creativity 387, 388, 392, 393
atmosphere of nursery school 17–18
attentiveness and cognitive development 365
attitude formation and social class 148
attitude transmission 138
attitudes and maturational status 192, 194, 195, 197, 198–9

Barker Lunn, J.C. 384, 386, 390, 391
Barron, F. 336–7, 380
Bennett, S.N. 105–6
Berger, L. 64–5
Bernstein, B. 153
Biddle, W.E. 63
black IQ deficit 345–56
  as result of environmental *v.* genetic factors 345–6, 348–56
  as result of motivational deficit 353
  as result of teacher expectancy 353
  as result of test situation factors 353
  as result of verbal deprivation 352–3
Blacky Pictures 60, 64, 66, 69–71
blind children's development of articulated cognitive functioning 312–13
Block, W.E. 69–70
body concept
  and global *v.* articulated cognitive style 295, 300–1, 303
  development 305
Bronfenbrenner, U. 143, 144
browsing conditions and assessment of competence 258–9
Burt, C. 341

case study method 6
case study of treatment of underachiever 163–6
castration anxiety 61–3, 65–8, 69–70
  measurement 61–2
castration complex 44, 58–60, 61–3, 64–5, 66, 68, 70
casual observation 5–6
cathexis of psychical material 50, 51
Cattell, R.B. 78–89, 90–4, 96–8, 99–100, 341
chance, as viewed by different social classes 153
Child, I.L. 67
child-rearing procedures
  as antecedents of castration anxiety 67–8
  research 143–4
childhood sexuality 42–4
civilization environmental theory of intelligence 357, 365–6
civilization of group, as average level of cognitive function used in problem-solving 365–6
clash of values approach 143, 148
cluster analysis and classification procedures 104–5
cognitive competence 257–86
  measurement 258–60
cognitive mapping approach to understanding cognitive functioning

307–9, 310, 311
cognitive structures used in child's environment, and cognitive development 364–5
cognitive style 257–86, 288–312, 382
  and issues in education 288, 305–14
  and personality adjustment 288, 302
  as way of perceiving both world and self 288
  development 302–5
  in terms of cognitive competence 257–86
  predisposing to creativity 382
Cohen, E. 148
college and university level research on personality and achievement 84–5, 101–3
Community-Integrated programme of preschool education 168–70, 175–87
competence achievement, as process involving many small steps 159, 162
competence models 138–9
complexive grouping structures 209, 210, 217
Concept Mastery Test 336, 337
concept reproduction hypothesis 262
condensation in dreams 52–3
conditioned drive or anxiety 117–18, 123
conscious processes 47–55
consciousness 37, 45
  localization 37
conservation experiments 237, 238
constructs 19–32
  held by teachers of pupils 21–32
content analysis 259, 282–3, 284
  of teachback protocols 282–3
control groups 10, 248–51
controlled observation 8–12
controlled variables 9, 10
Cookson, D. 98–9
correlational analysis 6–8, 81, 91–3, 103–5, 320–1
  of factors in 16PF 91–3
correspondence
  and equivalence of sets 222–36
  as means of comparing sets 221
Corwin, S.M. 65–7
Creative Analysis materials 266, 267–8, 271
  clobbits 267–8, 271
creative individuals' personal qualities 379–82
creativity
  and achievement 338–9, 340
  and curriculum structure 385–93
  and educational environment 384
  and informal/formal schools 386, 387–8, 390–3
  and intelligence or abilities 336–40, 368, 385–6, 388, 391–3
  and personality 379
  as criterion based 336–7
  as dependent on relationship of personality to curriculum structure 387
  as dependent on size and quality of the association reservoir 387
  as goodness of response 369
  as originality of response 368, 370
  as what creativity tests measure 337–9
  evaluation 368–78
  of architects 336
  of research workers 336
  of writers 337
  theories 382, 386–8, 392–3
criterion measured by tests 88
cross-cultural study
  of encouragement given to competence drives 137
  of equivalence formation 219
  of extent of field dependence 310
  of menstrual taboo 67–8
  of post-partum sexual taboo 69
cross-sectional studies 302, 303–4
cultural bias of IQ tests 349–51, 352
culture fair tests of intelligence 350, 353
Cunningham, S. 98–9
curiosity
  as a drive 119
  as an intrinsic motive 133
curriculum structure and creativity 385–93

Dacey, J. 385
Davis, A. 143, 146, 148–9
day residues in dreams 52–3
deaf children's IQ test performance 353
death or destructive instinct 40
degree of structure in preschool education programmes 168–73, 186
denial of feeling, and maturity status 193, 200
dependent variables 10
development
  of cognitive functioning 302–3, 312–13, 363–7
  of correspondence 221, 222–36
  of differentiation 303–5, 311–12
  of equivalence 204–19
  of invariance principle 242–53
Differential Aptitude Tests 332–4
differentiation of construct system 28, 30–2
disadvantaged children
  and preschool education 168, 169, 173, 177–8, 187–8
  learning with advantaged children 177–8, 187
displacement in dreams 53
Dollard, J. 118
dominance and achievement 79
Douvan, E. 151–2
dreams 51–5, 377
drives, as motivators of behaviour 117
Duvall, E.M. 153
dynamic tension levels and achievement 82–3

educational environment and creativity 384
educational implications of research
  on cognitive competence 257
  on cognitive style 288, 305–14
  on impulsivity/reflexivity 326
educational materials for serialists *v.* holists 260, 277
ego 38–9, 41, 44, 49–52, 54–5
  preconscious nature 49–50
  and dreams 51–2, 54–5
Embedded Figures Test 291, 313
enactive systems of representation 204
environment
  and race 358, 359, 362, 366
  as cumulative 360, 361, 362–3
  as determinant of abstract cognitive structures 357, 363–7
  as determinant of all cognitive skills 357
  as series of spurts 361, 362, 363
  between family and within family variation 358
environmental enrichment and relationship of race to IQ 361, 363
environmental *v.* genetic explanations of intelligence 345–6, 348–56, 357–67
equality of opportunity 355
equivalence formation 204–19, 222–36
equivalence of sets 221–36
  and correspondence judgments 222–36
Eros or love/preservation instinct 40
erotogenic zones 41, 43
  as a source of libido 41
errors on inductive reasoning test 315, 317, 318–19, 321–5
evaluation
  in education 306–10
  of creativity 368–78
  of intellectual performance 368–70
  of preschool education programmes 168–88
  of research papers 15–16
examinations, and teachback testing 257, 262
Experimental programme of preschool education 168–9, 171–3, 175–88
Extrapolation Reasoning Test 319, 322
extraversion/introversion
  and performance 78–80, 97–103, 105–6
  as a construct system 23, 28, 30
  as related to Oral Scale 76
  measurement problems 106–7
extrinsic learning motives 141
exvia-invia factor of personality 90, 92
eye scanning in visual discrimination tasks 315, 317, 324–5
Eysenck, H.J. 76, 90–4, 96–9, 101–3, 106–8

factor analysis 330–7, 340
factor analytic positions 78–89, 90–4, 96–110
factored tests 89
family environment, as a linear variable 154
family relationships, and development of differentiation 303–5, 311–12
fear
  as conditioned response 123–4
  as response to familiar objects in strange surroundings 124–5
fear of death, and castration anxiety 65–6
Festinger, L. 126
fiat mode of defining equivalence 206, 207
field dependence *v.* independence in perception 292–4, 300, 305, 313–14
  and ability 293
  and analytical IQ 293
  and mode of cognitive functioning 300
  and problem-solving performance 293
  and sex 313–14
  development 305
Figure Drawing Technique in study of body concept 295, 300, 303
fixations of the libido 44
flexibility of creative individuals 380–1
focused observation 6–8
form of research papers 13–14
formal/informal schools
  and creativity 386, 387–8, 390–3
  as dichotomy 386
  as linear 386
Fraser, E. 142–3, 144
free learning conditions, and assessment of competence 258–9, 271–6, 284
Freud, S. 37–55, 56–71, 73–7, 117, 118, 122, 377
Friedman, S.M. 57–60
Fromm, E. 381
Frostig, M. 174, 184, 185
Frostig Developmental Test of Visual Perception 174, 184, 185
functional mode of defining equivalence 206, 207–9, 215–16

general intelligence 329–32, 337, 340, 341, 342
  and creativity 337
  as process 342
general intelligence tests and multitest batteries 333–4
genetic *v.* environmental explanations of intelligence 345–6, 348–56, 357–67
genital phase of sexual development 44
Getzels, J.W. 338, 389, 392
Ghiselli, E.E. 335
gifted children identification 339–40
global *v.* articulated cognitive style 294–5, 300–3, 305, 312–13
  and body concept 295, 300–1
  and sense of separate identity 301
  and view of self in different contexts 301
  as stages of development 302–3, 313
  development 302–3, 305, 312–13
Golann, S.E. 368
Gottheil, E. 73–7
grouping in classroom
  implications of research on cognitive style 310
grouping principles employed
  and age 209–12, 217–18
Guessing Objects Test 319, 322
Guilford, J.P. 369, 370, 379, 384, 389

habit, as explanation for behaviour 118–19
  habit systems 102–3
Haddon, P.A. 384, 385, 386, 390, 391, 392, 393
Haller, A.O. 146
Hammer, E. 60
Hammer, E.F. 65
Haptic Visual Matching Test 317, 321–3
Harrell, T.H. 335
Hayes, K.J. 341
'Headstart' environmental theory of intelligence 357, 363, 366
Hebb, D.O. 121, 124–6, 133, 341

Helson, H. 126
Henmon-Nelson Test 338
heritability of intelligence 349, 357
higher education sector
  and relationship of personality to achievement 102
Himmelweit, H.T. 148–9
holists and holistic learning strategy 258, 259, 260, 266, 268, 271, 275–8, 279–80, 281–6
  and educational materials 260, 277
  and nature of predicates used in hypothesis testing 259, 272, 284
homosexuality 44, 62–3
  and the castration complex 62–3
Hull, C.L. 115, 118, 120
Hunt, J.McV. 1, 116–27, 341
Hyman, H.H. 149–50
hypercathexis of psychical material 50
hypnosis, as means of investigating Oedipus complex 63
hypothesis conflict theory and children's quantity judgments 237, 238, 251

id 38–9, 41, 49–53, 55
  unconscious nature 49–50
  and dreams 51–3, 55
  illogical nature 53
identification processes 137–8
  self sustaining nature 137
ideographic method 35
ikonic systems of representation 204, 219
Illinois Test of Psycholinguistic Abilities 171, 174, 178–83, 185, 188
impulsivity/reflection and inductive reasoning 315–26
incongruity or dissonance theory of fear, and motivation 124–7
inconsistency of response in conservation experiments
  as resolution of conflict between incompatible hypotheses 237, 238, 251
  as result of lack of understanding of effect of perceptual transformation 237
independent variables 10
Indians' intelligence test performance 350, 354
individual differences 342–3
inductive reasoning, and reflection/impulsivity 315–26
informal/formal schools
  and creativity 386, 387–8, 390–3
  as dichotomy 386
  as linear 386
information processing 127
initiation rites 69
instincts 39–42
integrated *v.* segregated schooling and relation of race to IQ 362
integration of construct system 28, 30–2
intelligence or abilities
  and achievement 78, 81, 84, 107–8, 300, 346–7, 351
  and creativity 336–40, 368–70, 385–6, 388, 391–3
  and effectiveness of preschool intervention 174, 175
  and fear 123, 124
  and field dependence/independence 293
  and interests 84
  and preschool education 174–7, 185, 187
  and race 345–56, 361–3, 366
  and relationship of personality to achievement 78, 80, 81, 84, 107–8
  and relationship of social class to aspirations 146
  and test situation factors 352
  as capacity to abstract 365
  as correctness of response 369
  as layer-cake phenomenon 363–4
  as process 342
  as set of abilities 307–9
  as unitary phenomenon 294, 300, 329–32, 341, 342
  measurement 333–4
  of managers 355
  of parents and children 359, 361, 365
  of rural *v.* urban people 366–7
  regression effects 358–9

interaction approach to study of mother-child relations 303
interests 84
internalization of values, and social class 151–3
interpersonal attitudes and maturational status 192, 194, 195–6, 198–9
interpretation of dreams 51–5
intervening variables in relationship of personality to achievement 107–10
intervention programmes for disadvantaged preschool children 168–88
intrinsic learning motives 132–41
invariance principle 237, 243, 245, 246, 252
Israeli boys' psychosexual development 64
item analysis of Oral Scale 73–6
item difficulty of IQ tests
  and race 351
  and nationality 351
  and sex 351
item intercorrelations of IQ tests, and race 352

Jackson, P.W. 338, 368–83, 389, 392
Jacob, F. 265, 267
Jensen, A.R. 345–56, 357–66

Kahl, J.A. 150–1
Katz, F.M. 152–3
Kelly, G.A. 17, 19, 127
kibbutz *v.* family rearing, and psychosexual development 64
kinship data on intelligence 357–8, 362
Kites–Fishes Puzzle Design Task 389
Kluckhohn, H. 69
Kogan, H. 389
Kuppuswamy, B. 56

laboratory setting 12
language development and preschool education 174, 178–83, 185, 187–8
language frames and modes of defining equivalence 206
language of tester and IQ test performance 352
Lasky, J.L. 64–5
latency period 43
latent content of dreams 51, 53–4
learning environment and relationship of personality to achievement 84–5, 107–9
learning strategies 257–86
learning theory view of treatment of underachiever 167
learning *v.* thinking as educational objectives 340
level of cognitive functioning of institutions 366
libido 40–1, 44
life orientations and aspirations 148
Line Meanings Task 389
longitudinal studies 302–4
Lytton, H. 384, 385, 386, 390, 391, 392, 393

McClelland, D.C. 126, 137
McElroy, W.A. 56–7
Mackinnon, D.W. 336, 392
Maccoby, E. 313
Madaus, G. 385
malnutrition and mental development 355
manifest content of dreams 51, 53–4
match/mismatch of learning strategy and teaching material 260, 262, 266, 277, 278, 279–82, 285, 286
matched samples 9, 147, 385, 386
  in study of social class and aspiration 147
  in study of creativity 385, 386
Matching Familiar Figures Test 315, 317, 321–4
maturational status and personality 192–203
Mednik, S.A. 69, 70, 387
memory/rote learning *v.* abstract reasoning tasks, and performance of different races 353
menstrual taboos and the castration complex 67–8
Mexican Americans' intelligence test performance 350, 354
Michal-Smith, H. 60
Miller, N.E. 118

minority group performance on IQ tests 350
mobility and social class value systems 149–50
mobility pessimism of father and achievement 153–4
modes of defining equivalence 206–9, 215–17
Monod, J. 265, 267
Montessori programme of preschool education 8–71, 175–87
mother-child relationship, and development of differentiation 303–5, 311–12
motivated *v.* spontaneous behaviour 117, 119–20
motivation
  and achievement 78, 80–5
  and experience 116–27
  and maturational status 192, 193, 199
  and parental encouragement 145–6
  and personality 144–5
  and situational factors 144–5
  and social class value systems 143, 146–55
  as determinant of black IQ deficit 353
  as intrinsic *v.* extrinsic to information processing 127
  in terms of the incongruity-dissonance principle 125–7
  theories 116–25
motivational distortion of test responses 87
motivational-experiential theory of intelligence 341
motives 132–41
multitest battery validity 332–4
Murray, A. 146

national group differences in item difficulty on IQ tests 351
natural environment 12
naturalistic observation 6
need for achievement, and maturity status 193, 199
neuroticism/stability
  and achievement 97–103
  measurement problems 106–7
Newman–Keuls Tests 175
nominal mode of defining equivalence 206, 207, 216
nomothetic approach 35
Number Series Task 389
nursery schools 17–18
nursery teachers' perceptions of pupils 17, 22

objective criteria of evaluation 369
observation as a source of information 4–12
occupational choice and cognitive style 311
Oedipus complex 44, 56–61, 63–4, 68, 69
Oedipus phase of sexual development 44
operational definition of variables 7
'Operon' controlled biochemical systems test 267, 286
optical *v.* operational correspondence 234
oral phase of sexual development 43, 73
Oral Trait Scale validity 73–6
orality 73–6
originality of creative individuals 380
others' observations as a source of information 12

parallel forms of 16PF 91
parental dominance as perceived by son, and maturational status 192, 195
parental encouragement, and achievement 145–6
parental influence
  and adolescent development 202–3
  and relation of ego to superego 39
  and child's cognitive development 365
parental rejection as perceived by son,
  and maturational status 192–3, 195, 197
parental value systems research 143–55
pathology and global *v.* articulated cognitive style 302
Patterson, G.R. 163
Peabody Picture Vocabulary Test 174, 184, 185, 351, 352

Pearson product moment coefficient 7–8
peer group
  as means for overcoming underachievement 163–7
  as language model 177–8
perceptible mode of defining equivalence 206, 207, 215–16, 217
perception
  in terms of field dependence/independence 292
  of the upright in space 289–91
perceptual transformations and children's quantity judgments 222–36, 237–52
personal observations as a source of information 5–12
personal qualities of creative individuals 379–82
personality
  and ability 78, 80, 81, 84, 107–8
  and achievement 78–85, 96–110
  and cognitive style 288, 302
  and creativity 379–82
  and interests 84
  and maturational status 192–203
  and the environment 144–5
  measurement 73–6, 78–89, 90–4, 96–8, 106–8, 192
  types 104
perversions 44
phallic phase of sexual development 43–4
philosophers view of consciousness 45–7
philosophers' view of consciousness
Piaget, J. 119, 126, 221–36, 237, 238, 251, 252, 364
pictorial materials and equivalence formation 213–18
Picture Completion Reasoning Test 318, 322
Pidgeon, D.A, 385
placement in education
  implications of research on cognitive style 306, 310–11
poetic qualities of creative individuals 381–2
political *v.* scientific explanations of black IQ deficit 346, 356
post-partum sexual taboo 69
preconscious processes 47–9, 51–2, 54
preschool education 17, 168–88
primary factors
  in intelligence 330–1
  in personality measurement 90–4, 97, 99
problem-solving behaviour and reflection/impulsivity 315–26
process nature of intelligence 342
projective methods in study of Oedipus complex 58–60, 61–3, 64–7, 69–70
psyche or mental life 37
psychical qualities 45–50
  as unconscious 45
psycho-analytic theory 37–55, 56–71, 73–7, 377
psychological test use 86–9

quantity judgments in children 221–36, 237–52
questioning *v.* projective methods in study of Oedipus complex 58–60

Rabin, A.I. 64
race
  and environment 358, 359, 361, 362, 366
  and intelligence or abilities 345–6, 361–3, 366
  and interpersonal environment 361
  and regression effect for IQ 358–9
race of tester and IQ test performance 352
racism 346
random sampling 9
rank-order coefficient 7–8
Ravens Progressive Matrices 355
reading research papers 13–16
reciprocity as an intrinsic learning motive 139–40
redundant *v.* irredundant holists 258, 259, 275, 277
  and educational materials 277
  and question distribution amongst card-category co-ordinates 259

reference group
  and relationship of social class and aspirations 148
  identification 137
reflection/impulsivity and inductive reasoning 315–26
regression 45
regression effect for IQ, and race 358–61
reinforcement 120, 121, 127, 132–41, 158–60, 162, 163–7, 388
  and stimulus variation 121
  as contingent on accomplishment of learning steps 159, 162
  as intrinsic *v.* extrinsic to information processing 127
  for underachievers 163–7
  frequency in the classroom 159–60, 162
  internal *v.* external forms 132–41
  of learning 158–60
reliability
  of extraversion and neuroticism scales 106–7
  of TAT categories 194
Remote Association Test 369–70
Repertory Grid Technique 17, 19
research design effects on research results 103–8
residence area and relationship of personality to achievement 107
response time on inductive reasoning tests 315, 317, 318–19, 321–5
retarded children's cognitive functioning 308–9
retarded *v.* accelerated physical development in adolescence, and personality 194–203
review of research
  on creativity and educational environment 384–6
  on Oedipus and castration complexes 56–71
  on personality and achievement 98–103
  on social class and achievement 142–55
Rod-and-Frame Test 290–1
Rogers, C.R. 126
Roheim, G. 60–1
role of nursery school teacher 17
Rosen, B.C. 150–1
Rudin, S.A. 301
sadistic-anal phase of sexual development 43
sampling, as a means of variable control 9
Sampson, L.W. 266
Sarnoff, I. 65–7
satisfaction, as aesthetic response to creative product 374, 375, 377, 378
savouring, as aesthetic response to creative product 378
school-level research on personality and achievement 84–5, 98–101
school placement
  implications of research on cognitive style 310
school psychological service
  implications of research on personality and achievement 85–9
school readiness and preschool education 174, 185
school's inefficiency in teaching drill subjects 160
school's role
  in channelling curiosity motive 134
  in channelling competence motive 136
Schwartz, B.J. 61–3
scientific *v.* political explanations of black IQ deficit 346, 356
second order factors
  in intelligence 330–1
  in personality measurement 90–4, 97, 100
segregrated *v.* integrated schooling and relation of race to IQ 362
self conceptions and maturational status 192, 200–1
self evaluation of test performance 319–20
semantics of equivalence establishment 204–5, 206–9, 212
sense of separate identity

  and global *v.* articulated cognitive style 301
  and social context 301
  development 304, 305
sensitivity of creative individuals 380
sentential groupings 217–18
serialists and serial learning strategy 258, 259, 260, 266, 268, 271, 272, 275–8, 280–2, 284–6
  and educational materials 260, 277
  and nature of predicates used in hypothesis testing 259, 272, 284
Sewell, W.H. 146
sex differences
  in body adjustment tests 290
  in castration anxiety 62–3
  in cognitive functioning 310, 313
  in embedded figures tests 292
  in field dependence 290, 291, 292, 310, 314
  in problem solving behaviour 321–3
  in Rod-and-Frame Tests 291
  in item difficulty on IQ tests 351
sexual development 42–5
situational *v.* personality factors in motivation 144–5
16PF battery 90–4
social aspects of psychological testing 87–9
social class
  and achievement 142–55
  and aspirations 146–8
  and fear 123–4, 125
  and values 146–52
social orientation of underachievers 163
socially desirable test responses 87
sociological view of social class and achievement 143–55
'sociologists fallacy' 354
Spearman, C. 330, 341
Spitz, H. 60
stability
  of cognitive functioning 302–3
  of scores on extraversion and neuroticism scales 106–7
stages of development
  of correspondence 222–36
  of equivalence formation 206–12, 215–18
Stagner, R. 301
Stanford Binet Intelligence Scale 174, 175, 185, 187, 329, 332, 338, 352
Stephens, W.N. 67–8
stimulation, as aesthetic response to creative product 373, 375, 377, 378
stimuli of behaviour 117–25
  positive and negative exciting forms 121
Stone, G.C. 73–7
strategy change 265
study habits and relationship of personality to achievement 103
subject matter as source of reinforcement 161–2
subject performance and personality 79–80
subjective criteria of evaluation 369
success definitions, and social class 152–3
succorance and nurturance need, and maturity status 193, 199
summary power of product
  as judgmental standard for evaluating creativity 375–7
  and aesthetic response of savouring 378
  and poetic qualities of creative individual 381–2
super-ego 39, 41
  as preconscious 49
superordinate grouping structures 209–10, 218
surprise, as aesthetic response to creative product 374, 375, 377, 378
Symbol Series Task 389
symbolic systems of representation 204, 219
syntax of equivalence establishment 204–5, 209–12
system learning tasks 267, 286
systems of representation 204, 219

Taylor, D.W. 336
taxonomy learning tasks 266, 267–8, 271, 276
teachback
  and examinations 257, 262

  - as a learning procedure 257, 259–60, 262–3, 277, 278, 285, 286
  - order analysis 283–4
- teacher, as a competence model 139
- teacher cognitive characteristics, and effectiveness of pupil learning 311
- teacher expectancy, as explanation for black IQ deficit 353
- teacher personality, and relationship of pupil personality to achievement 108
- teacher-pupil interaction and effectiveness of preschool education 186
- teachers' constructs of pupils 21–32
- teacher's role in development of differentiation in child 311–12
- teaching methods and strategies
  - and relationship of personality to achievement 107–10
  - for serialists *v.* holists 260, 276–7
  - implications of research on cognitive style 306, 311–14
- temporal characteristics of eye scanning 315, 317, 324–5
- test development 88–9, 306–7
- test situation influence on IQ test performance 352, 361
- Thematic Apperception Test 192–200, 301, 304, 305
- thematic grouping structures 209, 211–12
- theories and explanations
  - of black IQ deficit 345–6, 348–56
  - of competence achievement 159, 162
  - of creativity 382, 386–8, 392–3
  - of environment 357, 360–7
  - of fear 123–5
  - of intelligence 329–32, 341, 342, 363–4, 365, 369
  - of motivation and learning 116–25
  - of response in conservation experiments 237, 238, 251
- threshold theory of relationship of curriculum structure to creativity 388, 392–3
- Thurstone, L.L. 330, 336
- time-interval study 6
- Torrance, E.P. 384–5, 392
- Traditional programme of preschool education 168–70, 175–87
- training children in reliability of correspondence cues and understanding of invariance 246–51, 252
- trait *v.* state anxiety 102
- transformation of unconscious into consciousness 48–9, 51–2
- transformation power of product
  - as judgmental standard for evaluating creativity 372–3, 375, 377
  - and aesthetic response of stimulation 373, 375, 378
  - and flexibility as property of creative individual 380–1
- treatment programme for underachievers 163–7
- two groups design 11
- type of school and relationship of personality to achievement 85, 107

- unconscious processes 47–55
- underachievers
  - social orientation 163
  - treatment programme 163–7
- uniqueness in relation to norms of product
  - as judgmental standard for evaluating creativity 370–1, 372–3, 375, 377
  - and aesthetic response of surprise 374, 375, 378
  - and originality as property of creative individual 380
- unitary nature of intelligence 294, 300, 307–9
- Upton, A. 266
- Uses Task 389

- validity
  - of creativity tests 336, 337, 340
  - of Differential Aptitude Tests 333
  - of IQ tests 351
  - of multitest batteries 332–3
  - of Oral Scale 73–6
  - of tests for different races 351
- value systems

  and achievement 143, 146–55
  and mobility 149–50
  and motivation 143, 146–55
  as learning environments 146
  of different social classes 146–52
value transmission 138, 148–9
  and social class 148–9
variables 7, 9,10
Ventur, P.A. 69–70
verbal deprivation as explanation for black IQ deficit 352–3
verbal materials and equivalence formation 205–13
verbal misunderstanding as explanation of mistakes in correspondence judgments 225–6, 228–9
Vernon, P.E. 384, 385
Veroff, J. 144–5
visual perception and preschool education 174, 184–5
vocabulary comprehension and preschool education 174, 183–4

Waetjen, W.B. 313
Wallach, L.M. 389
Warburton, F.W. 99–101
Weschler Intelligence Scales for Children 293–4, 313, 317, 319, 322, 332, 338, 350
Whiting, J.W.M. 67, 69
widespread *v.* implementary values 150–1
will to learn, as an intrinsic learning motive 132
wish fulfillment and dreams 54–5

zonal analysis 103–4

If we are to shape education more closely to the needs of individuals, we must understand better how differences in personality, considered in its widest sense, can produce different reactions in the learning situation. The two volumes of *Personality and Learning*, by drawing together two important areas of study which are more often considered separately, make a significant contribution to such an understanding.

These volumes contain articles which have been specially selected to accompany the Open University course *Personality and Learning* (E201). Each volume is divided into sections which examine the interaction between personality and learning from a number of different standpoints and in a variety of contexts. In this volume, a general introductory section offers useful guidelines for students who are approaching educational psychology for the first time. The reader then goes on to consider psychoanalytic and some factor-analytic theories of personality, learning and motivation, the course of development, cognitive styles, and intelligence and its relationship with creativity. The second volume looks at the topics of social relationships, cultural influences on cognition and attainment, adjustment problems, learning difficulties and classroom interaction. Together, the two volumes provide a broadly based introduction to educational and child psychology which will be invaluable to all those with a serious interest in psychology and education.

£3.75
net in UK

ISBN 0 340 20472 9
Hodder and Stoughton
in association with The Open University Press